ASYMPTOTIC VARIANCE FOR RANDOM WALK METROPOLIS CHAINS IN HIGH DIMENSIONS: LOGARITHMIC GROWTH VIA THE POISSON EQUATION

ALEKSANDAR MIJATOVIĆ AND JURE VOGRINC

Abstract. There are two ways of speeding up MCMC algorithms: (1) construct more complex samplers that use gradient and higher order information about the target and (2) design a control variate to reduce the asymptotic variance. While the efficiency of (1) as a function of dimension has been studied extensively, this paper provides the first results linking the efficiency of (2) with dimension. Specifically, we construct a control variate for a d-dimensional Random walk Metropolis chain with an IID target using the solution of the Poisson equation for the scaling limit in [RGG97]. We prove that the asymptotic variance of the corresponding estimator is bounded above by a multiple of $\log(d)/d$ over the spectral gap of the chain. The proof hinges on large deviations theory, optimal Young's inequality and Berry-Esseen type bounds. Extensions of the result to non-product targets are discussed.

1. Introduction

Markov Chain Monte Carlo (MCMC) methods are designed to approximate expectations of high dimensional random vectors, see e.g. [BGJM11, Tie94]. It is hence important to understand how the efficiency of MCMC algorithms scales with dimension. The optimal scaling literature, initiated by the seminal paper [RGG97], indicates that for the high-dimensional algorithms it is the growth of the asymptotic variance with dimension that provides perhaps the most natural measure of efficiency for MCMC (see [RR01, Sections 1.2 and 2.2], [RR98, Sec. 3]). For instance, for a product target, the asymptotic variance for the Random walk Metropolis (RWM) chain on $\mathbb{R}^d$ is heuristically of the order $\mathcal{O}(d)$ [RGG97, RR01]. Moreover, the asymptotic variances of the d-dimensional Metropolis-adjusted Langevin algorithm (MALA), Hamiltonian Monte Carlo and the fast-MALA are $\mathcal{O}(d^{1/3})$ [RR98], $\mathcal{O}(d^{1/4})$ [BPR$^+$13] and $\mathcal{O}(d^{1/5})$ [DRVZ16], respectively.

This paper constructs a dimension-dependent estimator (see (1) below) and proves a bound on its asymptotic variance, suggesting the order $\mathcal{O}(\log(d))$, for a RWM chain with an IID target. The idea is to exploit the following facts: (I) the law of the diffusion scaling limit for the RWM

Date: March 28, 2019.

2000 *Mathematics Subject Classification.* 60J10, 60J22.

Key words and phrases. Asymptotic variance, Poisson equation for diffusions, variance reduction, scaling limits of Metropolis-Hastings chains, optimal Young's inequality, Berry-Esseen bounds, large deviations.

We thank Krys Latuszynski for insightful comments. AM is supported in part by the EPSRC grant EP/P003818/1 and the Programme on Data-Centric Engineering funded by Lloyd's Register Foundation. JV is supported by President's PhD Scholarship of Imperial College London and by the EPSRC grants EP/P002625/1 and EP/R022100/1.

chain (as $d \to \infty$) from [RGG97] is close (in the weak sense) to that of the law of the chain itself *and* (II) the Poisson equation for the limiting Langevin diffusion has an explicit solution. Following ideas from [MV16], we construct and analyse the estimator using (I) and (II).

Specifically, let ρ be a density on $\mathbb{R}$ and $\rho_d(\mathbf{x}^d) := \prod_{i=1}^d \rho(x_i^d)$ the corresponding d-dimensional product density, where $\mathbf{x}^d = (x_1^d, \dots, x_d^d) \in \mathbb{R}^d$. Let $\mathbf{X}^d = \{\mathbf{X}_n^d\}_{n\in\mathbb{N}}$ be the RWM chain converging to ρ_d with the normal proposal with variance $l^2/d \cdot I_d$ (here $I_d \mathbf{x}^d = \mathbf{x}^d$, for all $\mathbf{x}^d \in \mathbb{R}^d$, and l a constant), analysed in [RGG97]. If $f(\mathbf{x}^d)$ depends only on the first coordinate x_1^d, then $\rho_d(f) := \int_{\mathbb{R}^d} f(\mathbf{x}^d)\rho_d(\mathbf{x}^d)d\mathbf{x}^d = \int_{\mathbb{R}} f(x)\rho(x)dx =: \rho(f)$. Under appropriate conditions, the asymptotic variance $\sigma^2_{f,d}$ in the Central limit theorem (CLT) for the estimator $\sum_{i=1}^n f(\mathbf{X}_i^d)/n$ of $\rho_d(f)$ satisfies

$$\sigma^2_{f,d} \leqslant 2\mathrm{Var}_\rho(f)/(1-\lambda_d) \qquad \text{and, heuristically,} \qquad \sigma^2_{f,d} = \mathcal{O}(d) \text{ as } d \to \infty.$$

Here $\mathrm{Var}_\rho(f) := \rho(f^2) - \rho(f)^2$ and $1-\lambda_d$ denotes the spectral gap of the chain $\mathbf{X}^d$. The inequality follows by the spectral representation of $\sigma^2_{f,d}$ in [Gey92, KV86]. The reasoning analogous to that applied to the integrated autocorrelation time in [RR01, Sec. 2.2] can be used to argue that the spectral gap $1-\lambda_d$ is of the order $\mathcal{O}(1/d)$. Hence the asymptotic variance $\sigma^2_{f,d}$ is $\mathcal{O}(d)$.

The Poisson equation for the Langevin diffusion arising in the scaling limit of $\mathbf{X}^d$ in [RGG97] is a second order linear ODE with solution $\hat{f}$ given explicitly in terms of f and the density ρ, see (7) below. For large d, the function $\hat{f}$ ought to approximate the solution of the Poisson equation for $\mathbf{X}^d$. The reasoning in [MV16] then suggests the form of an estimator for $\rho_d(f)$, which under appropriate technical assumptions, satisfies the following CLT:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \left(f + d(P_d\hat{f} - \hat{f})\right)(\mathbf{X}_i^d) - \rho_d(f)\right) \xrightarrow{d} N(0, \hat{\sigma}^2_{f,d}) \quad \text{as } n \to \infty, \tag{1}$$

where P_d is the transition kernel of the chain $\mathbf{X}^d$. The main result of this paper (Theorem 1 below) states that for some constant $C > 0$ the following inequality holds

$$\hat{\sigma}^2_{f,d} \leqslant C\log(d)/(d(1-\lambda_d)) \qquad \text{and, heuristically,} \qquad \hat{\sigma}^2_{f,d} = \mathcal{O}(\log(d)) \text{ as } d \to \infty.$$

Theorem 1 also gives the explicit dependence of the constant C on the function f.

This result suggests that to achieve the same level of accuracy as when estimating $\rho_d(f)$ by an average over an IID sample form ρ_d, only $\mathcal{O}(\log d)$ times as many RWM samples are needed if the control variate $d(P_d\hat{f} - \hat{f})$ is added. This should be contrasted to $\mathcal{O}(d)$ (resp. $\mathcal{O}(d^{1/3})$, $\mathcal{O}(d^{1/4})$, $\mathcal{O}(d^{1/5})$) times as many samples for the RWM (resp. MALA, Hamiltonian Monte Carlo, fast-MALA) without the control variate, see [RR01, RR98, BPR+13, DRVZ16].

The optimal scaling for the proposal variance of a d-dimensional RWM chain is $\mathcal{O}(1/d)$, see [RR01] for a review and [BRS09, Thm. 4] for the proof that other scalings lead to suboptimal behaviour. To get a non-trivial scaling limit in [RGG97], it is necessary to accelerate the chains $(\mathbf{X}^d)_{d\in\mathbb{N}}$ linearly in dimension. The weak convergence of the accelerated chain to the Langeivn diffusion suggests that $d\hat{f}$ is close to the solution of the Poisson equation for P_d and f, making $d(P_d\hat{f} - \hat{f})$ a good control variate. Using an approximate solution to the Poisson equation to construct control variates is a common variance reduction technique, see e.g. [Hen97, Mey08, DK12, MV16]. In this context, more often than not, an approximate solution used is a solution of a Poisson equation of a simpler related process. For instance, in [MV16] a sequence of Markov chains on a finite state space, converging weakly to a given RWM chain, is constructed. Then a solutions to the Poisson equation of the finite state chain is used to construct a control variate capable of reducing the asymptotic variance of the RWM chain arbitrarily. Here this idea is turned on its head: a solution to the Poisson equation of the limiting diffusion is used to construct a control variate for a RWM chain from a weakly convergent sequence in [RGG97]. Since the complexity of the RWM increases arbitrarily as dimension $d \to \infty$, it is infeasible to get an arbitrary variance reduction as in [MV16]. However, heuristically, the amount of variance reduction measured by the ratio $\sigma^2_{f,d}/\hat{\sigma}^2_{f,d}$ still tends to infinity at the rate $d/\log(d)$.

If the solution of the Poisson equation for $\mathbf{X}^d$ and f were available, we could construct an estimator for $\rho_d(f)$ with zero variance (see e.g. [MV16]). Put differently, in this case there would be no need for the chain to explore its state space at all. In our setting, since the jumps of $\mathbf{X}^d$ are of size $\mathcal{O}(1/\sqrt{d})$ [RGG97], after $\mathcal{O}(\log d)$ steps the chain will have explored the distance of $\mathcal{O}(\log(d)/\sqrt{d})$. In line with the observation above this distance tends to zero as $d \to \infty$ since, heuristically, $d\hat{f}$ approximates the solution of the Poisson equation for P_d and f arbitrarily well.

The key technical step in the proof of our result (Theorem 3 below) is a type of concentration inequality. It generalises the limit in [RGG97, Lemma 2.6], which essentially states that generators of the accelerated chains $(\mathbf{X}^d)_{d\in\mathbb{N}}$ converge to the generator of the Langevin limit when applied to a compactly supported and infinitely smooth function, in two ways: (A) it extends the limit to a class of functions of sub-exponential growth and (B) provides estimates for the rate of convergence. Both of these extensions are key for our main result. (A) allows us to apply Theorem 3 to a solution of the Poisson equation, which is not compactly supported. Note that this step in the proof entails identifying the correct space of functions that is closed under the operation of solving the Poisson equation (see Proposition 21 below). Estimate (B) allows us to control the asymptotic variance via a classical spectral-gap bound. The proof of Theorem 3, outlined in Sec. 4.1 below, crucially depends on the large deviations theory (Sec. 5.2), the form of the constant in optimal Young's inequality (Sec. 5.3) and Berry-Esseen type bounds (Sec. 5.4).

We conclude the introduction with a comment on how the present paper fits into the literature. Since, as discussed above, the asymptotic variance $\sigma^2_{f,d}$ is approximately equal to the product $2\mathrm{Var}_\rho(f)/(1-\lambda_d)$, two "orthogonal" approaches to speeding up MCMC algorithms are feasible. (a) The MCMC method itself can be modified, with the aim of increasing the spectral gap, leading to many well-known reversible samplers such as MALA and Hamiltonian Monte Carlo [DKPR87] as well as non-reversible ones [BR17, DLP16]. There is a plethora of papers (see [RR98, BPR+13, DRVZ16] and the references therein) studying the asymptotic properties of such sampling algorithms as dimension increases to infinity. (b) A control variate g, satisfying $\rho(g) = 0$, may be added to f with the aim of reducing $\mathrm{Var}_\rho(f)$ to $\mathrm{Var}_\rho(f+g)$ without modifying the MCMC algorithm. A number of control variates have been proposed in the MCMC literature [AC99, PMG14, OGC17, DK12]. Thematically, the present paper fits under (b) and, to the best of our knowledge, is the first to investigate the growth of the asymptotic variance as the dimension $d \to \infty$ in this context. Moreover, it is feasible that our method could be generalised to some of the algorithms under (a), see Section 3.2 below for a discussion of possible extensions.

The remainder of paper is structured as follows. Section 2 gives a detailed description of the assumptions and states the results. Section 3 illustrates algorithms based on our main result with numerical examples and discusses (without proof) potential extensions of our results for other MCMC methods, more general targets, etc. In Section 4 we prove our results. Section 5 develops the tools needed for the proofs of Section 4. Section 5 uses results from probability and analysis but is independent of all that precedes it in the paper.

2. Results

Let $\mathbf{X}^d = \{\mathbf{X}^d_n\}_{n\in\mathbb{N}}$ be a RWM chain in $\mathbb{R}^d$ with a transition kernel $P_d f := (1/d)\mathcal{G}_d f + f$, where

$$\mathcal{G}_d f(\mathbf{x}^d) := d\mathbb{E}_{\mathbf{Y}^d}\left[\left(f(\mathbf{Y}^d) - f(\mathbf{x}^d)\right)\alpha(\mathbf{x}^d, \mathbf{Y}^d)\right], \qquad \alpha(\mathbf{x}^d, \mathbf{Y}^d) := 1 \wedge \frac{\rho_d(\mathbf{Y}^d)}{\rho_d(\mathbf{x}^d)}, \tag{2}$$

$\mathbf{x}^d \in \mathbb{R}^d$, $\mathbf{Y}^d = (Y_1^d, \ldots, Y_d^d) \sim N(\mathbf{x}^d, l^2/d \cdot I_d)$ and $x \wedge y := \min\{x, y\}$ for all $x, y \in \mathbb{R}$, started in stationarity $\mathbf{X}^d_1 \sim \rho_d$. Let $\mathcal{S}^n$ consists of all the functions with their first n derivatives growing slower then any exponential function. More precisely, for any $n \in \mathbb{N} \cup \{0\}$, define

$$\mathcal{S}^n := \left\{ g \in \mathcal{C}^n(\mathbb{R}) \colon \sum_{i=0}^{n} \|g^{(i)}\|_{\infty,s} < \infty \ \forall s > 0 \right\}, \qquad \text{where } \|g\|_{\infty,s} := \sup_{x\in\mathbb{R}} \left(e^{-s|x|}|g(x)|\right) \tag{3}$$

and $\mathcal{C}^n(\mathbb{R})$ (resp. $\mathcal{C}^0(\mathbb{R})$) denotes n-times continuously differentiable (resp. continuous) functions. Our main result (Theorem 1 below) applies to the space $\mathcal{S}^1$, containing functions f for which $\rho(f) := \int_{\mathbb{R}} f(x)\rho(x)dx$ is typically of interest in applications (e.g. polynomials). In addition, spaces in (3) are closed for solving Poisson's equation in (6), see Proposition 21 below.

Throughout the paper ρ denotes a strictly positive density on $\mathbb{R}$ with $\log(\rho) \in \mathcal{S}^4$ and

$$\lim_{|x|\to\infty} \frac{x}{|x|} \cdot \log(\rho(x))' = -\infty, \tag{4}$$

unless otherwise stated. Assumption (4) implies that the tails of ρ decay faster then any exponential, i.e. $\mathbb{E}\left[e^{sX}\right] < \infty$ for any $s \in \mathbb{R}$ for $X \sim \rho$ (cf. [JH00, Sec. 4]). The assumption $\log(\rho) \in \mathcal{S}^4$ prohibits ρ from decaying to quickly, e.g. proportionally to $e^{-e^{|x|}}$. Both of these assumptions serve brevity and clarity of the proofs and it is feasible they can be relaxed. Nevertheless, a large class of densities of interest satisfy these assumptions, e.g. mixtures of Gaussian densities or any density proportional to $e^{-p(x)}$ for a positive polynomial p.

The scaling limit, introduced in [RGG97], of the chain $\mathbf{X}^d$ as the dimension d tends to infinity is key for all that follows. Consider a continuous-time process $\{U_t^d\}_{t\geqslant 0}$, given by $U_t^d := X^d_{\lfloor d\cdot t\rfloor,1}$, where $\lfloor\cdot\rfloor$ is the integer-part function and $X^d_{\cdot,1}$ is the first coordinate of $\mathbf{X}^d$ (since the proposal distribution for $X^d_{\cdot,1}$ has variance l^2/d, time needs to be accelerated to get a non-trivial limit).

As shown in [RGG97] (see also [RR01]), the weak convergence $U^d \Rightarrow U$ holds as $d \uparrow \infty$, where U is the Langevin diffusion started in stationarity, $U_0 \sim \rho$, with generator acting on $f \in \mathcal{C}^2(\mathbb{R})$ as

$$\mathcal{G}f := (h(l)/2)(f'' + (\log\rho)'f'), \quad \text{where } h(l) := 2l^2\Phi\left(-l\sqrt{J}/2\right) \text{ and } J := \rho\left((\log(\rho)')^2\right) \tag{5}$$

and Φ is the distribution of $N(0,1)$. Poisson's equation for U and a function f takes the form

$$\mathcal{G}\hat{f}(x) = \rho(f) - f(x). \tag{6}$$

It is immediate that a solution $\hat{f}$ of (6) is given by the formula

$$\hat{f}(x) := \int_0^x \frac{2dy}{h(l)\rho(y)} \int_{-\infty}^y \rho(z)(\rho(f) - f(z))dz, \qquad x \in \mathbb{R}. \tag{7}$$

In the remainder of the paper $\hat{f}$ denotes the particular solution in (7) of the equation in (6).

As usual, for $p \in [1,\infty)$, $f : \mathbb{R}^d \to \mathbb{R}$ is in $L^p(\rho_d)$ if and only if $\|f\|_p := \rho_d(|f|^p)^{1/p} < \infty$. Finally, note that under our assumptions on ρ, the kernel P_d of the RWM chain $\mathbf{X}^d$ defined above is a self-adjoint bounded operator on the Hilbert space $\{g \in L^2(\rho_d) \colon \rho_d(g) = 0\}$ with norm $\lambda_d < 1$

Theorem 1. *If $f \in \mathcal{S}^1$, then $\hat{f} \in \mathcal{S}^3$ and CLT (1) holds for the function $f + dP_d\hat{f} - d\hat{f}$ and the RWM chain $\mathbf{X}^d$ introduced above. Furthermore, there exists a constant $C_1 > 0$, such that for all $f \in \mathcal{S}^1$ and $d \in \mathbb{N}\backslash\{1\}$, the asymptotic variance $\hat{\sigma}^2_{f,d}$ in CLT (1) satisfies:*

$$\hat{\sigma}^2_{f,d} \leqslant C_1 \left(\sum_{i=1}^3 \|\hat{f}^{(i)}\|_{\infty,1/2}\right)^2 \frac{\log(d)}{(1-\lambda_d)d}.$$

The proof of Theorem 1 is given in Section 4.4 below. It is based on the spectral-gap estimate of the asymptotic variance $\hat{\sigma}^2_{f,d} \leqslant 2\|\mathcal{G}_d\hat{f} - \mathcal{G}\hat{f}\|_2^2/(1-\lambda_d)$ from [Gey92, KV86], the uniform ergodicity of the chain $\mathbf{X}^d$ and the following proposition.

Proposition 2. *There exists a constant C_2 such that for every $f \in \mathcal{S}^3$ and all $d \in \mathbb{N}\backslash\{1\}$ we have: $\|\mathcal{G}f - \mathcal{G}_d f\|_2 \leqslant C_2\left(\sum_{i=1}^3 \|f^{(i)}\|_{\infty,1/2}\right)\sqrt{\log(d)/d}$.*

The proof of Proposition 2, given in Section 4.3 below, requires a pointwise control of the difference $\mathcal{G}f - \mathcal{G}_d f$ on a large subset of $\mathbb{R}^d$. To formulate this precisely, we need the following.

Definition 1. A positive sequence $a = \{a_d\}_{d\in\mathbb{N}}$ is **sluggish** if the following holds:

$$\lim_{d\to\infty} a_d = \infty \quad \text{and} \quad \sup_{d\in\mathbb{N}\backslash\{1\}} \frac{a_d}{\sqrt{\log d}} < \infty.$$

Theorem 3 below is the main technical result of the paper. It generalises the limit in [RGG97, Lemma 2.6] to a class of unbounded functions and provides an error estimate for it. The bound in Theorem 3 yields sufficient control of the difference $\mathcal{G}f - \mathcal{G}_d f$ to establish Proposition 2.

Theorem 3. *Let $a = \{a_d\}_{d\in\mathbb{N}}$ be a sluggish sequence. There exist constants $c_3, C_3 > 0$ (dependent on a) and measurable sets $\mathcal{A}_d \subset \mathbb{R}^d$, such that for all $d \in \mathbb{N}$ we have $\rho_d(\mathbb{R}^d\backslash\mathcal{A}_d) \leqslant c_3 e^{-a_d^2}$ and*

$$\left|\mathcal{G}f(x_1^d) - \mathcal{G}_d f(\mathbf{x}^d)\right| \leqslant C_3\left(\sum_{i=1}^3 \|f^{(i)}\|_{\infty,1/2}\right) e^{|x_1^d|}\frac{a_d}{\sqrt{d}} \qquad \text{for any } f \in \mathcal{S}^3 \text{ and } \mathbf{x}^d \in \mathcal{A}_d.$$

The proof of Theorem 3 is outlined and given in Sections 4.1 and 4.2 below, respectively.

Remark 1. The dependence on f in the bound of Theorem 3 is not sharp. The factor $\sum_{i=1}^{3}\|f^{(i)}\|_{\infty,1/2}$ is used because it states concisely that the speed of the convergence of $\mathcal{G}_d f$ to $\mathcal{G}f$ depends linearly on the first three derivatives of f. Moreover, it is not clear if the bound in Theorem 1 and Proposition 2 is optimal in d. However, if an improvement were possible, the proof would have to be significantly different to the one presented here. In particular, a better control of the difference $|\mathcal{G}f-\mathcal{G}_d f|$ on $\mathbb{R}^d\backslash\mathcal{A}_d$ would be required.

3. Discussion and numerical examples

3.1. **Discussion.** In this section we discuss potential extensions of Theorem 1 to settings satisfying weaker assumptions or involving related MCMC chains.

3.1.1. *IID target for non-RWM chains in stationarity.* Perhaps the most natural generalisation of Theorem 1 would be to the MALA and fast-MALA chains (in [RR98] and [DRVZ16]) for which it is also possible to obtain non-trivial weak Langevin diffusion limits under appropriate scaling. Since the form of the Poisson equation in (6) is preserved it is possible to define an estimator like the one in (1) and it seems feasible that a version of Theorem 3 can be established in this context using methods analogous to the ones in this paper.

3.1.2. *IID target in the transient phase for the RWM chain.* Theorem 1 is a result only about the stationary behaviour of the chain. As in practice MCMC chains are typically started away from stationarity it is important to understand the transient behaviour. In [JLM15] it is shown that the scaling limit described in Section 2 above has mean-field behaviour of the McKean type, i.e. the limiting process is a continuous semimartingale with characteristics that at time t depend on the law of the process at t. This suggests that an appropriately chosen time-dependent function $\hat{f}$ in the estimator in (1) could further reduce the constant in the bound of Theorem 1.

3.1.3. *General product target density.* The class of target distributions considered in [Béd07, BR08], preserves the independence (i.e. product) structure but allows for a different, dimension dependent, scaling of each of the components of the target law. If the proposal variances appropriately reflect the scaling in the target, each component in the infinite dimensional limit is a Langevin diffusion. Again, as in Section 3.1.1 above, the estimator in (1) can be applied directly and an extension of Theorem 1 to this setting appears feasible.

3.1.4. *Gaussian targets in high dimensions.* Let π_0 denote a Gaussian target on $\mathbb{R}^d$ with mean μ and covariance matrix $\mathrm{diag}(\sigma_{11}^2,\dots,\sigma_{dd}^2)$. Inspired by [RR01, Thm 5] and the proof of Theorem 1, a good control variate for the ergodic average estimator for $\pi_0(f)$ takes the form $d(P_d\tilde{f}-\tilde{f})$, where $\tilde{f}$ solves the ODE

$$\tilde{f}''+\Big((\partial/\partial x_1^d)\log\pi_0\Big)\tilde{f}'=2/h_0(l)\cdot(\pi_0(f)-f),$$

with $h_0(l) := 2l^2\Phi(-l\sqrt{J_0}/2)$ and $J_0 := 1/d \cdot \sum_{j=2}^d 1/\sigma_{jj}^2$. In the case of the mean, $f(x_1^d) = x_1^d$, we can solve the ODE explicitly: $\tilde{f}(x_1^d) = 2\sigma_{11}^2/h(l) \cdot x_1^d$.

If π_0 has a general non-degenerate covariance matrix Σ, we have an ODE analogous to the one above for each eigen-direction of Σ. The control variate for the mean of the first coordinate, $f(x_1^d) = x_1^d$, is then a linear combination of the control variates for the means in eigen-directions. Specifically, $\tilde{f} : \mathbb{R}^d \to \mathbb{R}$ in the estimator analogous to the one in (1) (see the numerical example for $h = 0$ in Section 3.2.2) takes the form

$$(8)\quad \tilde{f}(\mathbf{x}^d) = 2/h_0(l) \cdot \sum_{j=1}^{d} x_j^d \Sigma_{j1}, \quad \text{with } h_0(l) = 2l^2\Phi(-l\sqrt{J_0}/2) \text{ and } J_0 = 1/d \cdot \mathrm{Tr}((\Sigma^{-1})_{2:d,2:d}).$$

Note that $\tilde{f}$ does not depend on the mean of the target, a special feature of the Gaussian setting.

3.1.5. *Non-product target density.* A typical non-product target density considered in the literature [BRS09, MPS12, PST12] is a projection of a probability measure Π on a separable real Hilbert space $\mathcal{H}$ onto a d-dimensional subspace, where Π is given via its Radon-Nikodym derivative $\frac{d\Pi}{d\Pi_0}(x) \propto \exp(-\Psi(x))$. Here Ψ is a densely defined positive functional on $\mathcal{H}$ and Π_0 is a Gaussian measure on $\mathcal{H}$ specified via a positive trace-class operator on $\mathcal{H}$ (see e.g. [MPS12, Sec.2.1] for a detailed description and [BRS09] for the motivation for this class of measures). A key feature of this framework is that there exists an $\mathcal{H}$-valued Langevin diffusion z, driven by a cylindrical Brownian motion on $\mathcal{H}$ (i.e. a solution of an SPDE), that describes the scaling limit of the appropriately accelerated sequence of chains $(\mathbf{X}^d)_{d\in\mathbb{N}}$, see [MPS12, PST12].

An estimator analogous to the one in (1) would require the solution $\hat{f}$ of the Poisson equation of z. While this might be a feasible strategy theoretically, it would likely be difficult to numerically evaluate the solution $\hat{f}$. However, inspired by the Simplified Langevin Algorithm in [BRS09], which uses as the proposal chain an Euler scheme for the Langevin diffusion with target Π_0 (not Π), we suggest constructing the estimator for $\Pi(f)$ in (1), with $\hat{f}$ the solution of the Poisson equation for the Langevin diffusion converging to Π_0 (not Π). This strategy is feasible as we are able to produce good control variates for Gaussian product targets in high dimensions in the spirit of Theorem 1, see Section 3.1.4 above.

The main theoretical question in this context is to find suitable assumptions on the functional Ψ in the Radon-Nikodym derivative above that guarantee the asymptotic variance reduction as $d \to \infty$. Expecting an improvement from polynomial to logarithmic growth is unrealistic as we are solving the Poisson equation for Π_0 instead of Π. However, the numerical example in Section 3.2.2 suggests that this idea may work in practice if Π is close from Gaussian Π_0. Understanding in which settings does it lead to significant variance reduction is another relevant and important question.

3.2. **Numerical examples.** The basic message of the present paper is that the process in the scaling limit of an MCMC algorithm contains useful information that can be utilised to achieve significant savings in high dimensions.

In both examples presented below the problem is to estimate the mean of the first coordinate $\rho_d(f)$, where $f(\mathbf{x}^d) := x_1^d$. A run of T steps of a well-tuned RWM algorithm with kernel P_d, defined in the beginning of Section 2, started in stationarity, produces a RWM sample $\{\mathbf{X}_n^d\}_{n=1,2,\cdots T}$ and an estimate $\hat{\rho}_d(f) := \sum_{n=1}^T f(\mathbf{X}_n^d)/T$ of $\rho_d(f)$. Take $\tilde{f}$ to be the associated solution of the Poisson equation, that we obtain numerically in the first example 3.2.1 and using formula (8) in the second example 3.2.2. In both cases we estimate the required unknown quantities ($\rho_d(f)$ and Σ) from the sample $\{\mathbf{X}_n^d\}_{n=1,2,\cdots T}$.

Using the same sample, define $\tilde{\rho}_d(f) := \frac{1}{T}\sum_{n=1}^T (f + dP_d\tilde{f} - d\tilde{f})(\mathbf{X}_n^d)$. Since the function $P_d\tilde{f} - \tilde{f}$ is not accessible in closed form, for every $n \leqslant T$ we use IID Monte Carlo to estimate the value $(P_d\tilde{f} - \tilde{f})(\mathbf{X}_n^d)$ as $\sum_{j=1}^{n_{MC}} \left(1 \wedge (\rho_d(\mathbf{Y}_n^{d,j})/\rho_d(\mathbf{X}_n^d))\right)\left(\tilde{f}(\mathbf{Y}_n^{d,j}) - \tilde{f}(\mathbf{X}_n^d)\right)/n_{MC}$, where $\mathbf{Y}_n^{d,1}, \ldots, \mathbf{Y}_n^{d,n_{MC}}$ is an IID sample of size n_{MC} from $N(\mathbf{X}_n^d, l^2/d \cdot I_d)$. This estimation step can be parallelised (i.e. run on n_{MC} cores simultaneously).

We measure the variance reduction due to the post processing above by comparing the mean square errors of $\hat{\rho}_d(f)$ and $\tilde{\rho}_d(f)$ as estimators of $\rho_d(f)$ over n_R independent runs of the RWM chain,

$$\mathrm{VR}(\rho, f) := \frac{\sum_{k=1}^{n_R}((\hat{\rho}_d(f))_k - \rho_d(f))^2}{\sum_{k=1}^{n_R}((\tilde{\rho}_d(f))_k - \rho_d(f))^2}, \tag{9}$$

where $(\hat{\rho}_d(f))_k$ and $(\tilde{\rho}_d(f))_k$ are the averages in the k-th run of the chain. Heuristically, this means the estimator $\tilde{\rho}_d(f)$ of $\rho_d(f)$ based on T sample points is as good as estimator $\hat{\rho}_d(f)$ based on $\mathrm{VR}(\rho, f) \cdot T$ sample points.

3.2.1. *Multi-modal product target.* To verify that what theory predicts also happens in practice we first present an example of an IID target with each coordinate a bimodal mixture of two Gaussian densities. Given the results in Table 1, we wish to highlight the robustness of the method with respect to numerically estimating $\tilde{f}$ and $(P_d\tilde{f} - \tilde{f})(\mathbf{X}_n^d)$ for each $n \leqslant T$.

Let ρ be a mixture of two normal densities $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, with the first arising in the mixture with probability $2/5$ and $\mu_1 = -3, \mu_2 = 4$ and $\sigma_1 = \sigma_2 = 7/4$. The potential of the density ρ has two wells and is in a well know class arising in models of molecular dynamics, see e.g. [DLP16, Sec. 5.4]. Note that the corresponding density ρ_d on $\mathbb{R}^d$, defined in the introduction, has 2^d modes.

With variance reduction (9) we measure how much the estimator $\tilde{\rho}_d(f)$ outperforms the estimator $\hat{\rho}_d(f)$ of the mean of the first coordinate $\rho_d(f) = 6/5$. The results across a range of dimensions d and values of the parameter n_{MC} (that corresponds to the accuracy of the estimation of $P_d\tilde{f} - \tilde{f}$) are presented in Table 1. All the entries were computed using $n_R = 500$ independent runs of length $T = 2 \cdot 10^5$.

To numerically solve the Poisson equation in (6), substitute the derivatives of f and $\log(\rho)$ with symmetric finite differences and use the estimate $\hat{\rho}(f)$ for $\rho(f)$. Recall that the standard deviation of the proposal in our RWM algorithm is $l/\sqrt{d}$. The solver uses a grid of hundred points equally spaced in the interval $[\min_{n\leqslant T} \mathbf{X}_{n,1}^d - 3 \cdot l/\sqrt{d}, \max_{n\leqslant T} \mathbf{X}_{n,1} + 3 \cdot l/\sqrt{d}]$, where

$d \backslash n_{MC}$	30	50	70	150	300
5	17.5	20.5	23.9	28.4	28.9
10	20.9	36.5	40.4	54.7	78.9
20	31.4	46.4	65.0	105.6	116.4
30	35.5	51.3	71.0	127.1	163.1
50	35.7	55.7	77.8	136.8	196.8

TABLE 1. Variance reduction for different dimensions d and values of n_{MC}.

$\mathbf{X}^d_{n,1}$ is the first coordinate of the n-th sample point. We use the Moore-Penrose pseudoinverse as the linear system is not of full rank. Finally, we take $\tilde{f}$ to be the linear interpolation of the solution on the grid. Note that this is a crude approximation of the solution of (6), which does not exploit analytical properties of either f or ρ.

The results in Table 1 contain a lot of noise due to numerically solving the ODE, using an approximation $\hat{\rho}_d(f)$ for $\rho_d(f)$ and using IID Monte Carlo to estimate $\tilde{\rho}_d(f)$. It is interesting to note, that despite these additional sources of error, the variance reduction is considerable and behaves as the theoretical results predict. The estimator $\tilde{\rho}_d(f)$ improves with dimension, and with n_{MC}. Increasing n_{MC}, however, has diminishing effect which is particularly clear in the case $d = 5$. Due to the asymptotic nature of our result we can only expect limited gain for any fixed d, even if $P_d\tilde{f} - \tilde{f}$ could be evaluated exactly (corresponding to $n_{MC} = \infty$).

3.2.2. *Bi-modal non-product target.* Can the theoretical findings of this paper help us construct control variates in more realistic cases with non-product target densities? It is unreasonable to expect a simple general answer to this question. A more realistic approach for future work seems to be trying to establish specific forms of control variates that work well for classes of targets of certain type. We briefly explore one such instance in this section. Sections 3.1.4 and 3.1.5 as well as the results in Table 2 suggest we can construct useful control variates when the target is close to a Gaussian.

Let $\mu_{d,h}$ be a d-dimensional vector with entries $(h/2, 0, \ldots, 0)$ for $h \geqslant 0$ and let $\Sigma^{(d)}$ be a $d \times d$ covariance matrix with the largest eigenvalue equal to $\lambda = 25$ with the corresponding eigenvector being $(1, 1 \ldots 1)$ and all other eigenvalues being equal to one. Take $\Pi_{d,h}$ to be the mixture of two d-dimensional normal densities $N(-\mu_{d,h}, \Sigma^{(d)})$ and $N(\mu_{d,h}, \Sigma^{(d)})$, both arising in the mixture with probability 1/2.

We wish to estimate the mean of the first coordinate $\Pi_{d,h}(f) = 0$ (for $f(\mathbf{x}^d) = x_1^d$). To produce a control variate we simply pretend, that we are dealing with a Gaussian target instead of $\Pi_{d,h}$. Let $\Sigma^{\Pi_{d,h}}$ be the covariance of $\Pi_{d,h}$ and $\hat{\Sigma}^{\Pi_{d,h}}$ an estimate of $\Sigma^{\Pi_{d,h}}$ obtained from the RWM sample $\{\mathbf{X}^d_n\}_{n=1,2,\ldots,T}$. Define $\tilde{f}_{d,h}$ as in (8) using $\hat{\Sigma}^{\Pi_{d,h}}$. We compare the performance of estimators $\hat{\Pi}_{d,h}(f)$ and $\tilde{\Pi}_{d,h}(f)$ of $\Pi_{d,h}(f) = 0$, respectively defined as $1/T \sum_{n=1}^T f(\mathbf{X}^d_n)$ and $1/T \sum_{n=1}^T \left(f + dP_d\tilde{f}_{d,h} - d\tilde{f}_{d,h}\right)(\mathbf{X}^d_n)$, according to variance reduction (9).

Table 2 shows the results across a range of dimensions d and distances between modes h, which measures the 'non-Gaussianity' of the target. Note that when $h = 0$ the target is Gaussian $N(0, \Sigma^{(d)})$ which we include to demonstrate the validity of control variate (8) for Gaussian targets. All the entries were calculated using $n_R = 500$ independent runs of length $T = 2 \cdot 10^5$ and $n_{MC} = 50$ IID Monte Carlo steps for computing $P_d\tilde{f} - \tilde{f}$ at each time step.

$d\backslash h$	0	2	4	6	8	10
5	60.1	40.5	34.6	3.78	1.21	1.01
10	59.3	38.2	12.4	1.88	1.05	1.00
20	46.8	37.6	7.00	1.51	1.01	1.00
30	37.7	36.2	5.88	1.31	1.01	1.00
50	27.8	25.8	3.50	1.26	1.00	1.00

TABLE 2. Variance reduction for different dimensions d and distances between modes h.

The quality of results decays with dimension because the proposal is scaled as $1/d$ in each coordinate. This results in the first coordinate mixing slower and for $h \neq 0$ also being less able to cross between modes, hence our estimate $\hat{\Sigma}^{\Pi_{d,h}}$ of the covariance becomes worse as we are working with fixed RWM sample length T. When $h = 10$ (and some cases of $h = 8$) it is unlikely that the RWM sample will reach the other mode at all which results in no gain from the method.

If we use the true covariance $\Sigma^{\Pi_{d,h}}$ of the target in the control variate (8), instead of learning it from the sample $\hat{\Sigma}^{\Pi_{d,h}}$, the corresponding results for $d = 50$ are presented in Table 3.

h	0	2	4	6	8	10
$d = 50$	73.0	59.2	3.91	1.31	1.02	0.99

TABLE 3. Variance reduction for dimension $d = 50$ and different distances between modes h using the true covariance of the target.

Unsurprisingly the estimator $\tilde{\Pi}_{d,h}(f)$ does not perform well when the distance between modes h is large. Interestingly though, the method does offer considerable gain in cases $h = 2$ and $h = 4$, even a noticeable gain in $h = 6$. For $h = 4$ and $h = 6$ the target is already clearly bimodal and different from the Gaussian, the RWM sample stays in the same mode for hundreds, respectively thousands of time-steps at a time.

4. PROOFS

Throughout this section we assume the sluggish sequence $a = \{a_d\}_{d\in\mathbb{N}}$ is given and fixed and, as mentioned above, the density ρ satisfies $\log(\rho) \in \mathcal{S}^4$ and has sub-exponential tails (4). Section 4.1 outlines the proof of Theorem 3 by stating the sequence of results that are needed to establish it. The proofs of these results, given in Section 4.2, rely on the theory developed in Section 5 below. Sections 4.3 and 4.4 establish Proposition 2 and Theorem 1, respectively.

4.1. **Outline of the proof of Theorem 3.** We start by specifying sets $\mathcal{A}_d \subset \mathbb{R}^d$ that have large probability under ρ_d. We need the following fact.

Proposition 4. *There exists a constant $c_A > 0$, such that the following open subset of $\mathbb{R}$, $A := \{x \in \mathbb{R}; |\log(\rho)''(x)| < (\log(\rho)'(x))^2, 1/c_A < |\log(\rho)'(x)| < c_A\}$, satisfies $\rho(A) > 0$.*

Let A satisfy the conclusion of Proposition 4 and recall the notation $\rho(f) = \int_{\mathbb{R}} f(x)\rho(x)dx$ for any appropriate function $f : \mathbb{R} \to \mathbb{R}$. Recall $J = \rho(((\log\rho)')^2) = -\rho((\log\rho)'')$, where the equality follows from assumptions $\log(\rho) \in \mathcal{S}^4$ and (4). Define the sets $\mathcal{A}_d$ as follows.

Definition 2. Any $\mathbf{x}^d \in \mathbb{R}^d$ is in $\mathcal{A}_d$ if and only if the following four assumptions hold:

$$\frac{1}{d-1}\sum_{i=2}^{d} e^{|x_i^d|} < 2\int_{\mathbb{R}} e^{|x|}\rho(x)dx, \tag{10}$$

$$\frac{1}{d-1}\sum_{i=2}^{d} 1_A(x_i^d) > \frac{\rho(A)}{2}, \tag{11}$$

$$\frac{1}{d-1}\left|\sum_{i=2}^{d}\left(\log(\rho)'(x_i^d)\right)^2 - J\right| < \frac{a_d}{\sqrt{d}}\sqrt{3\int_{\mathbb{R}}\left((\log(\rho)'(x))^2 - J\right)^2\rho(x)dx}, \tag{12}$$

$$\frac{1}{d-1}\left|\sum_{i=2}^{d}\log(\rho)''(x_i^d) + J\right| < \frac{a_d}{\sqrt{d}}\sqrt{3\int_{\mathbb{R}}\left(\log(\rho)''(x) + J\right)^2\rho(x)dx}. \tag{13}$$

Remark 2. The precise form of the constants in Definition 2 is chosen purely for convenience. It is important that $\int_{\mathbb{R}} e^{|x|}\rho(x)dx < \infty$ by (4), $\rho(A) > 0$ by Proposition 4 and that the constants in (12)–(13) are in $(0,\infty)$. Moreover, for any $\mathbf{x}^d \in \mathcal{A}_d$ there are no restrictions on its first coordinate x_1^d and the sets $\mathcal{A}_d$ are typical in the following sense.

Proposition 5. *There exists a constant c_1, such that $\rho_d(\mathbb{R}^d\backslash\mathcal{A}_d) \leqslant c_1 e^{-a_d^2}$ for all $d \in \mathbb{N}$.*

Using the theory of large deviations and classical inequalities, the proof of the proposition bounds the probabilities of sets where each of the above four assumptions in Definition 2 fails (see Sections 4.2 and 5.2 below for details).

Pick any $f \in \mathcal{S}^3$ and express the generator $\mathcal{G}_d$, defined in (2), as follows:

$$\mathcal{G}_d f(\mathbf{x}^d) = d\cdot\mathbb{E}_{Y_1^d}\left[\left(f(\mathbf{Y}^d) - f(\mathbf{x}^d)\right)\mathbb{E}_{\mathbf{Y}^d-}\left[1\wedge\frac{\rho_d(\mathbf{Y}^d)}{\rho_d(\mathbf{x}^d)}\right]\right], \qquad \mathbf{x}^d \in \mathbb{R}^d,$$

where $\mathbb{E}_{\mathbf{Y}^d-}[\cdot]$ is the expectation with respect to all the coordinates of the proposal $\mathbf{Y}^d$ in $\mathbb{R}^d$, except the first one (identify $f \in L^1(\rho)$ with $f \in L^1(\rho_d)$ by ignoring the last $d-1$ coordinates). The strategy of the proof of Theorem 3 is to define a sequence of operators, "connecting" $\mathcal{G}_d$ and $\mathcal{G}$, such that each approximation can be controlled for $f \in \mathcal{S}^3$ and $\mathbf{x}^d \in \mathcal{A}_d$.

First, for $f \in \mathcal{S}^3$, define

$$\tilde{\mathcal{G}}_d f(\mathbf{x}^d) := d\cdot\mathbb{E}_{Y_1^d}\left[\left(f(Y_1^d) - f(x_1^d)\right)\beta(\mathbf{x}^d, Y_1^d)\right], \qquad \mathbf{x}^d \in \mathbb{R}^d, \tag{14}$$

where for any $y \in \mathbb{R}$,

$$\beta(\mathbf{x}^d, y) := \mathbb{E}_{\mathbf{Y}^{d-}}\left[1 \wedge \exp\left(\log(\rho)'(x_1^d)(y - x_1^d) + \sum_{i=2}^{d} K(x_i^d, Y_i^d)\right)\right], \qquad \mathbf{x}^d \in \mathbb{R}^d, \tag{15}$$

and for any $(x, y) \in \mathbb{R}^2$ we define

$$K(x, y) := \log(\rho)'(x)(y - x) + \frac{\log(\rho)''(x)}{2}(y - x)^2 + \frac{(\log(\rho)'(x))^3\, 1_A(x)}{3}(y - x)^3. \tag{16}$$

In (16), the set A satisfies the conclusion of Proposition 4 and the coefficient before $(y - x)^3$ is chosen so that it is uniformly bounded for all $x \in \mathbb{R}$. This property plays an important role in proving that we have uniform control over the supremum norms of certain densities, cf. Lemmas 10 and 11 below. We can now prove the following.

Proposition 6. *There exists a constant C, such that for every $f \in \mathcal{S}^3$ and all $d \in \mathbb{N}$ we have:*

$$\left|\mathcal{G}_d f(\mathbf{x}^d) - \tilde{\mathcal{G}}_d f(\mathbf{x}^d)\right| \leqslant C\|f'\|_{\infty,1/2} e^{|x_1^d|} d^{-1/2} \quad \forall \mathbf{x}^d \in \mathcal{A}_d.$$

The proof of Proposition 6 relies only on the elementary bounds from Section 5.1 below. The idea is to use the Taylor series of $\log(\rho)(Y_i)$ around x_i^d for every $i \in \{1, \dots, d\}$ and then prove that modifying terms of order higher then two if $i \in \{2, \dots, d\}$ (resp. one if $i = 1$) is inconsequential.

Define the operator $\hat{\mathcal{G}}_d f(\mathbf{x}^d)$ for any $f \in \mathcal{S}^3$ and $\mathbf{x}^d \in \mathbb{R}^d$ by

$$\begin{aligned} \hat{\mathcal{G}}_d f(\mathbf{x}^d) \;&:=\; \frac{l^2}{2} f''(x_1^d)\mathbb{E}_{\mathbf{Y}^{d-}}\left[1 \wedge e^{\sum_{i=2}^d K(x_i^d, Y_i^d)}\right] \\ &\quad+\; l^2 f'(x_1^d)\log(\rho)'(x_1^d)\mathbb{E}_{\mathbf{Y}^{d-}}\left[e^{\sum_{i=2}^d K(x_i^d, Y_i^d)} 1_{\left\{\sum_{i=2}^d K(x_i^d, Y_i^d) < 0\right\}}\right]. \end{aligned} \tag{17}$$

We can now prove the following fact.

Proposition 7. *There exists a constant C, such that for every $f \in \mathcal{S}^3$, and all $d \in \mathbb{N}$ we have:*

$$\left|\hat{\mathcal{G}}_d f(\mathbf{x}^d) - \tilde{\mathcal{G}}_d f(\mathbf{x}^d)\right| \leqslant C\left(\sum_{i=1}^{3} \|f^{(i)}\|_{\infty,1/2}\right) e^{|x_1^d|} d^{-1/2} \quad \forall \mathbf{x}^d \in \mathcal{A}_d.$$

Note that, if we freeze the coordinates $x_2^d, \dots, x_d^d$ in $\mathbf{x}^d$, the operator mapping $f \in \mathcal{S}^3$ to $x_1^d \mapsto \hat{\mathcal{G}}_d f(\mathbf{x}^d)$ generates a one-dimensional diffusion with coefficients of the same functional form as in $\mathcal{G}$, but with slightly modified parameter values. The proof of Proposition 7 is based on the third and second degree Taylor's expansion of $y \mapsto f(y)$ and $y \mapsto \beta(\mathbf{x}^d, y)$ (around x_1^d), respectively, applied to the definition of $\tilde{\mathcal{G}}_d$ in (14). The difficult part in proving that the remainder terms can be omitted consist of controlling $\frac{\partial^2}{\partial y^2}\beta(\mathbf{x}^d, y)$, as this entails bounding the supremum norm of the density of $\sum_{i=2}^d K(x_i^d, Y_i^d)$ uniformly in d. Condition (11), which forces a portion of the coordinates x_i^d of $\mathbf{x}^d$ to be in the set A where the densities of the corresponding summands $K(x_i^d, Y_i^d)$ can be controlled, was introduced for this purpose. The details, explained in Sections 4.2 and 5.3 below, rely crucially on the optimal version of Young's inequality.

Introduce the following normal random variable with mean $\mu_{\mathcal{N}}(\mathbf{x}^d) = \frac{l^2}{2d}\sum_{i=2}^d \log(\rho)''(x_i^d)$ and variance $\sigma^2_{\mathcal{N}}(\mathbf{x}^d) = \frac{l^2}{d}\sum_{i=2}^d \left(\log(\rho)'(x_i^d)\right)^2$:

$$\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d) := \frac{l^2}{2d}\sum_{i=2}^d \log(\rho)''(x_i^d) + \sum_{i=2}^d \log(\rho)'(x_i^d)(Y_i^d - x_i^d). \tag{18}$$

Define the operator $\breve{\mathcal{G}}_d f(\mathbf{x}^d)$ for $f \in \mathcal{S}^3$ and $\mathbf{x}^d \in \mathbb{R}^d$ by:

$$\begin{aligned} \breve{\mathcal{G}}_d f(\mathbf{x}^d) \quad &:= \quad \frac{l^2}{2} f''(x_1^d)\mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}\right] \\ &+ \quad l^2 f'(x_1^d)\log(\rho)'(x_1^d)\mathbb{E}_{\mathbf{Y}^d-}\left[e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}1_{\{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)<0\}}\right]. \end{aligned} \tag{19}$$

Proposition 8. *There exists a constant C, such that for every $f \in \mathcal{S}^3$ and all $d \in \mathbb{N}$ we have:*

$$\left|\breve{\mathcal{G}}_d f(\mathbf{x}^d) - \hat{\mathcal{G}}_d f(\mathbf{x}^d)\right| \leqslant C\left(\sum_{i=1}^2 \|f^{(i)}\|_{\infty,1/2}\right) e^{|x_1^d|} d^{-1/2} \quad \forall \mathbf{x}^d \in \mathcal{A}_d.$$

First we show that $|\mathbb{E}_{\mathbf{Y}^d-}[1 \wedge e^{\sum_{i=2}^d K(x_i^d,Y_i^d)}] - \mathbb{E}_{\mathbf{Y}^d-}[1 \wedge e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}]|$ is small (Lemma 13 below). Proving that $\mathbb{E}_{\mathbf{Y}^d-}\left[e^{\sum_{i=2}^d K(x_i^d,Y_i^d)}1_{\{\sum_{i=2}^d K(x_i^d,Y_i^d)<0\}}\right]$ and $\mathbb{E}_{\mathbf{Y}^d-}\left[e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}1_{\{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)<0\}}\right]$ are close is challenging, as it requires showing that the supremum norm of the difference between the distributions of $\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)$ and $\sum_{i=2}^d K(x_i^d, Y_i^d)$ decays as $d^{-1/2}$ uniformly in its argument. The proof of this fact mimics the proof of the Berry-Esseen theorem and relies on the closeness of the CFs (characteristic functions) of $\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)$ and $\sum_{i=2}^d K(x_i^d, Y_i^d)$. The particular form of $K(x, Y)$ makes it possible to explicitly calculate the CF of $K(x, Y)$, if $x \notin A$, and bound it appropriately, if $x \in A$. The details are explained in Sections 4.2 and 5.4 below.

Since $\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)$ is normal, it is possible to explicitly calculate the expectations $\mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}\right]$ and $\mathbb{E}_{\mathbf{Y}^d-}\left[e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}1_{\{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)<0\}}\right]$, see [RGG97, Prop. 2.4]. Using these formulae, Proposition 9, which implies Theorem 3, can be deduced from assumptions (12)–(13).

Proposition 9. *There exists a constant C, such that for every $f \in \mathcal{S}^3$ and all $d \in \mathbb{N}$ we have:*

$$\left|\mathcal{G}f(x_1^d) - \breve{\mathcal{G}}_d f(\mathbf{x}^d)\right| \leqslant C\left(\sum_{i=1}^2 \|f^{(i)}\|_{\infty,1/2}\right) e^{|x_1^d|}\frac{a_d}{\sqrt{d}} \quad \forall \mathbf{x}^d \in \mathcal{A}_d.$$

Remark 3. The bounds in Propositions 6–8 are of the order $\mathcal{O}(d^{-1/2})$. The order $\mathcal{O}(a_d/\sqrt{d})$ of the bound in Proposition 9 gives the order in the bound of Theorem 3.

4.2. **Proof of Theorem 3.**

Proof of Proposition 4. Let $\tilde{A} := \left\{x \in \mathbb{R}; |\log(\rho)''(x)| < (\log(\rho)'(x))^2\right\}$. It suffices to show that the open set $\tilde{A}$ is not empty, since $\tilde{A} = \cup_{n\in\mathbb{N}}(\tilde{A} \cap \{x \in \mathbb{R}; \frac{1}{n} < |\log(\rho)'(x)| < n\})$, so for some large n_0 the open set $\tilde{A} \cap \left\{x \in \mathbb{R}; \frac{1}{n_0} < |\log(\rho)'(x)| < n_0\right\}$ must have positive Lebesgue measure and we can take $c_A := n_0$.

Assume that $\tilde{A} = \varnothing$, i.e. $|u'| \geqslant u^2$ on $\mathbb{R}$, where $u := \log(\rho)'$. Since ρ satisfies (4), there exists $x_0 < 0$ and $C > 0$ such that $u > C$ on the interval $(-\infty, x_0)$. Moreover, since $|u'| \geqslant u^2 > C^2 > 0$, u' has no zeros on $(-\infty, x_0)$ and satisfies either $u' \geqslant u^2$ or $-u' \geqslant u^2$ on the half-infinite interval.

Since $(1/u)' = -u'/u^2$, integrating the inequalities $-u'/u^2 \leqslant -1$ or $-u'/u^2 \geqslant 1$ from any $x \in (-\infty, x_0)$ to x_0, we get $1/u(x_0) + x_0 - x \leqslant 1/u(x)$ and $1/u(x_0) + x - x_0 \geqslant 1/u(x)$. Since by assumption it holds $0 < 1/u < 1/C$ on $(-\infty, x_0)$, we get a contradiction in both cases. □

Proof of Proposition 5. Let B_1^d, B_2^d, B_3^d and B_4^d be the subsets of $\mathbb{R}^d$ where assumptions (10), (11), (12) and (13) are not satisfied, respectively. Note that $\mathbb{R}^d \backslash \mathcal{A}_d = B_1^d \cup B_2^d \cup B_3^d \cup B_4^d$.

Recall that by (4), the L'Hospital's rule implies $\lim_{|x|\to\infty} \frac{\log \rho(x)}{x} \to -\infty$ and hence $\rho(e^{s|x|}) < \infty$ for any $s > 0$. Since $\{a_d\}_{d\in\mathbb{N}}$ is sluggish, there exists $n \in \mathbb{N}$ such that $a_d \leqslant \sqrt{n \log d}$ for all $d \in \mathbb{N}$. Then, by Proposition 26 applied to functions $x \mapsto (e^{|x|} - \rho(e^{|x|}))/\rho(e^{|x|})$ and $x \mapsto 2(\rho(A) - 1_A(x))/\rho(A)$, respectively, there exist constants c_1', c_2' such that the inequalities $\rho(B_1^d) \leqslant c_1' d^{-n} \leqslant c_1' e^{-a_d^2}$ and $\rho(B_2^d) \leqslant c_2' d^{-n} \leqslant c_2' e^{-a_d^2}$ hold for all $d \in \mathbb{N}$.

Likewise, there exist constants c_3', c_4' such that $\rho(B_3^d) \leqslant c_3' e^{-a_d^2}$ and $\rho(B_4^d) \leqslant c_4' e^{-a_d^2}$. This follows by Proposition 24, applied to the sequence $\{a_d\}_{d\in\mathbb{N}}$ and functions $g_3(x) := (\log(\rho)'(x))^2 - J$ (with $t := \sqrt{3\rho(g_3^2)}$) and $g_4(x) := \log(\rho)''(x) + J$ (with $t := \sqrt{3\rho(g_4^2)}$), respectively. Hence $\rho(\mathbb{R}^d \backslash \mathcal{A}_d) \leqslant \rho(B_1^d) + \rho(B_2^d) + \rho(B_3^d) + \rho(B_4^d) \leqslant c_1 e^{-a_d^2}$ for $c_1 := \max\{c_1', c_2', c_3', c_4'\}$. □

Remark 4. The proof above shows that the subsets B_1^d and B_2^d are of negligible size in comparison to B_3^d and B_4^d, since the $n \in \mathbb{N}$ can be chosen arbitrarily large.

Proof of Proposition 6. Pick an arbitrary $\mathbf{x}^d \in \mathcal{A}_d$ and recall that $\alpha(\mathbf{x}^d, \mathbf{Y}^d)$ is defined in (2). Since $|1 \wedge e^x - 1 \wedge e^y| \leqslant |x - y|$ for all $x, y \in \mathbb{R}$, for every realization Y_1^d, Taylor's theorem implies

$$\left|\mathbb{E}_{\mathbf{Y}^d_-}[\alpha(\mathbf{x}^d, \mathbf{Y}^d)] - \beta(\mathbf{x}^d, Y_1^d)\right| \leqslant |\log(\rho)''(W_1^d)|(Y_1^d - x_1^d)^2 + T_1^d(\mathbf{x}^d) + T_2^d(\mathbf{x}^d), \tag{20}$$

where W_1^d satisfies $\log(\rho)''(W_1^d)(Y_1^d - x_1^d)^2/2 = \log(\rho)(Y_1^d) - \log(\rho)(x_1^d) - \log \rho'(x_1^d)(Y_1^d - x_1^d)$ and

$$\begin{aligned} T_1^d(\mathbf{x}^d) &:= \frac{1}{6}\mathbb{E}_{\mathbf{Y}^d_-}\left[\left|\sum_{i=2}^d \left(\log(\rho)'''(x_i^d) - 2(\log(\rho)'(x_i^d))^3 1_A(x_i^d)\right)(Y_i - x_i^d)^3\right|\right], \\ T_2^d(\mathbf{x}^d) &:= \frac{1}{24}\mathbb{E}_{\mathbf{Y}^d_-}\left[\sum_{i=2}^d |\log(\rho)^{(4)}(Z_i^d)|(Y_i - x_i^d)^4\right]. \end{aligned}$$

Here Z_i^d satisfies $\log(\rho)^{(4)}(Z_i^d)(Y_i^d - x_i^d)^4/4! = \log(\rho)(Y_i^d) - \sum_{j=0}^3 (\log \rho)^{(j)}(x_i^d)(Y_i^d - x_i^d)^j/j!$ for any $2 \leqslant i \leqslant d$. Recall $Y_i^d - x_i^d$ is normal $N(0, l^2/d)$, for some constant $l > 0$, and $\log(\rho) \in \mathcal{S}^4$. Hence we may apply Proposition 23 to the function $x \mapsto \log(\rho)'''(x) - 2(\log(\rho)'(x))^3 1_A(x)$ to get $T_1^d(\mathbf{x}^d) \leqslant C_1 \left(l^6/d^3 \sum_{i=2}^d e^{|x_i^d|}\right)^{1/2}$ for some constant $C_1 > 0$, independent of $\mathbf{x}^d$. Since $\mathbf{x}^d \in \mathcal{A}_d$, the assumption in (10) yields $T_1^d(\mathbf{x}^d) \leqslant C_1 l^3 (2\rho(e^{|x|}))^{1/2}/d$. Similarly, we apply Proposition 22 (with $f = \log \rho$, $n = k = 4$, $m = 1$, $s = 1$ and $\sigma^2 = l^2/d$) and assumption (10) to get $T_2^d(\mathbf{x}^d) \leqslant C_2 d^{-2} \sum_{i=2}^d e^{|x_i^d|} \leqslant C_2 d^{-1}$ for some constant $C_2 > 0$ and all $\mathbf{x}^d \in \mathcal{A}_d$.

Recall $f \in \mathcal{S}^3$ and let $\tilde{W}_1^d$ be as in Proposition 22, satisfying $f'(\tilde{W}_1^d)(Y_1^d - x_1^d) = f(Y_1^d) - f(x_1^d)$. Let $C > 0$ be such that $T_1^d(\mathbf{x}^d) + T_2^d(\mathbf{x}^d) \leqslant C d^{-1}$ for all $\mathbf{x}^d \in \mathcal{A}_d$. The bound in (20), Taylor's theorem applied to f and Cauchy's inequality yield:

$$\begin{aligned}
&\left|\mathcal{G}_d f(\mathbf{x}^d) - \tilde{\mathcal{G}}_d f(\mathbf{x}^d)\right| \leqslant d\mathbb{E}_{Y_1^d}\left[\left|f(Y_1^d) - f(x_1^d)\right| \left(|\log(\rho)''(W_1^d)|(Y_1^d - x_1^d)^2 + Cd^{-1}\right)\right] \\
&= d\mathbb{E}_{Y_1^d}\left[\left|f'(\tilde{W}_1^d)\log(\rho)''(W_1^d)(Y_1^d - x_1^d)^3\right|\right] + C\mathbb{E}_{Y_1^d}\left[\left|f'(\tilde{W}_1^d)(Y_1^d - x_1^d)\right|\right] \\
&\leqslant d\left(\mathbb{E}_{Y_1^d}\left[\left|f'(\tilde{W}_1^d)^2(Y_1^d - x_1^d)^3\right|\right]\mathbb{E}_{Y_1^d}\left[\left|\left(\log(\rho)''\right)^2(W_1^d)(Y_1^d - x_1^d)^3\right|\right]\right)^{1/2} \\
&\quad + C\mathbb{E}_{Y_1^d}\left[\left|f'(\tilde{W}_1^d)(Y_1^d - x_1^d)\right|\right] \leqslant \bar{C}d(\|f'\|^2_{\infty,1/2}e^{|x_1^d|}d^{-3/2}\cdot e^{|x_1^d|}d^{-3/2})^{1/2} + \bar{C}\|f'\|_{\infty,1/2}e^{|x_1^d|}d^{-1/2}.
\end{aligned}$$

The last inequality follows by three applications of Proposition 22, where $\bar{C} > 0$ is a constant that does not depend on f or $\mathbf{x}^d \in \mathcal{A}_d$. This concludes the proof of the proposition. □

Before tackling the proof of Proposition 7, we need the following three lemmas. Recall that $K(x,Y)$ is defined in (16) and the set A satisfies the conclusion of Proposition 4.

Lemma 10. *Pick $x \in A$ and let $Y \sim N(x, l^2/d)$ for some constant $l > 0$. Then $K(x,Y)$ has a density q_x satisfying $\|q_x\|_\infty \leqslant 4c_A\sqrt{d}/(3l\sqrt{2\pi})$.*

Proof. Existence of q_x follows from (16) and Proposition 28. By Proposition 4 we have $|\log(\rho)''(x)| < (\log(\rho)'(x))^2$ and $c_A > |\log(\rho)'(x)| > 1/c_A$. Consider the polynomial $y \mapsto p(y) := \log(\rho)'(x)y + \log(\rho)''(x)y^2/2 + (\log(\rho)'(x))^3 y^3/3$. By (16) it holds $p(Y - x) = K(x,Y)$. Since $p'(y) = \log(\rho)''(x)y + \log(\rho)'(x)(1 + \log(\rho)'(x)^2y^2)$, we have

$$\begin{aligned}
|p'(y)| &\geqslant |\log(\rho)'(x)|(1 + \log(\rho)'(x)^2y^2) - \left|\log(\rho)''(x)\right||y| \\
&> |\log(\rho)'(x)|(1 - |\log(\rho)'(x)y| + |\log(\rho)'(x)y|^2) > \frac{3}{4c_A},
\end{aligned}$$

where the second inequality holds since $|\log(\rho)''(x)| < (\log(\rho)'(x))^2$ and the third follows from $\inf_{z\in\mathbb{R}}\{1 - |z| + z^2\} = 3/4$ and $|\log(\rho)'(x)| > 1/c_A$. The lemma now follows by Proposition 29. □

Recall that the proposal is normal $\mathbf{Y}^d = (Y_1^d, \ldots, Y_d^d) \sim N(\mathbf{x}^d, l^2/d \cdot I_d)$.

Lemma 11. *For any $\mathbf{x}^d \in \mathcal{A}_d$, the sum $\sum_{k=2}^d K(x_i^d, Y_i^d)$ possesses a density $\mathbf{q}^d_{\mathbf{x}^d}$. Moreover, there exists a constant C_K such that $\|\mathbf{q}^d_{\mathbf{x}^d}\|_\infty \leqslant C_K$ holds for all $d \in \mathbb{N}$ and all $\mathbf{x}^d \in \mathcal{A}_d$.*

Proof. Fix $\mathbf{x}^d \in \mathcal{A}_d$ and, for each i, let q_i denote the density of $K(x_i^d, Y_i^d)$ as in the previous lemma. Since the components of $\mathbf{Y}^d$ are IID, we have $\mathbf{q}^d_{\mathbf{x}^d} = \ast_{i=2}^d q_i = q_A \ast q_{\mathbb{R}\backslash A}$, where $q_A := \ast_{x_i^d \in A} q_i$ and $q_{\mathbb{R}\backslash A} := \ast_{x_i^d \notin A} q_i$. By the definition of convolution and the fact that $q_{\mathbb{R}\backslash A}$ is a density, it follows that $\|\mathbf{q}^d_{\mathbf{x}^d}\|_\infty \leqslant \|q_A\|_\infty \|q_{\mathbb{R}\backslash A}\|_1 = \|q_A\|_\infty$. By Lemma 10 there exists $C > 0$ such that, for any $d \in \mathbb{N}$, it holds $\|q_i\|_\infty < C\sqrt{d}$ if $x_i^d \in A$. Condition (11) implies there are at least $(d-1)\rho(A)/2$ factors in the convolution $q_A = \ast_{x_i^d \in A} q_i$. Hence Proposition 27 applied to q_A yields $\|\mathbf{q}^d_{\mathbf{x}^d}\|_\infty \leqslant \|q_A\|_\infty \leqslant c\frac{C\sqrt{d}}{\sqrt{(d-1)\rho(A)/2}}$. This concludes the proof of the lemma. □

Lemma 12. *Let $\mathbf{x}^d \in \mathcal{A}_d$. The function $y \mapsto \beta(\mathbf{x}^d, y)$, defined in (15), is in $\mathcal{C}^2(\mathbb{R})$ and the following holds:*

(i) $0 < \beta(\mathbf{x}^d, y) \leqslant 1$ for all $y \in \mathbb{R}$;

(ii) $\beta(\mathbf{x}^d, x_1^d) = \mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\sum_{i=2}^d K(x_i^d, Y_i^d)}\right]$;

(iii) $\left|\frac{\partial}{\partial y}\beta(\mathbf{x}^d, y)\right| \leqslant \left|\log(\rho)'(x_1^d)\right|$ *for all* $y \in \mathbb{R}$;

(iv) $\frac{\partial}{\partial y}\beta(\mathbf{x}^d, x_1^d) = \log(\rho)'(x_1^d)\mathbb{E}_{\mathbf{Y}^d-}\left[e^{\sum_{i=2}^d K(x_i^d, Y_i^d)}1_{\left\{\sum_{i=2}^d K(x_i^d, Y_i^d) < 0\right\}}\right]$;

(v) $\left|\frac{\partial^2}{\partial y^2}\beta(\mathbf{x}^d, y)\right| \leqslant \left|\log(\rho)'(x_1^d)\right|^2 (C_K + 1)$ *for all* $y \in \mathbb{R}$ *and constant* C_K *from Lemma 11.*

Proof. (i) and (ii) follow from the definition in (15). Since $x \mapsto 1 \wedge e^x$ is Lipschitz (with Lipschitz constant 1) on $\mathbb{R}$, the family of functions $\{x \mapsto (1 \wedge e^{x+h} - 1 \wedge e^x)/h; h \in \mathbb{R}\backslash\{0\}\}$ is bounded by one and converges pointwise to $1_{\{x<0\}}e^x$ for all $x \in \mathbb{R}\backslash\{0\}$, as $h \to 0$. Hence the DCT implies that $\frac{\partial}{\partial y}\beta(\mathbf{x}^d, y)$ exists and can be expressed as

$$\log(\rho)'(x_1^d)\mathbb{E}_{\mathbf{Y}^d-}\left[e^{\log(\rho)'(x_1^d)(y-x_1^d)+\sum_{i=2}^d K(x_i^d, Y_i^d)}1_{\left\{\log(\rho)'(x_1^d)(y-x_1^d)+\sum_{i=2}^d K(x_i^d, Y_i^d)<0\right\}}\right], \tag{21}$$

implying (iii) and (iv). Let Φ_K^d denote the distribution of $\sum_{i=2}^d K(x_i^d, Y_i^d)$ and recall that by definition we have $e^x 1_{\{x<0\}} = 1 \wedge e^x - 1_{\{x \geqslant 0\}}$ for all $x \in \mathbb{R}$. Hence, by (21), it follows

$$\frac{\partial}{\partial y}\beta(\mathbf{x}^d, y) = \log(\rho)'(x_1^d)\left(\beta(\mathbf{x}^d, y) - 1 + \Phi_K^d\left(-\log(\rho)'(x_1^d)(y - x_1^d)\right)\right), \tag{22}$$

By Lemma 11, Φ_K^d is differentiable. Hence, by (22), $\frac{\partial^2}{\partial y^2}\beta(\mathbf{x}^d, y)$ also exists and takes the form:

$$\left(\log(\rho)'(x_1^d)\right)^2\left(\beta(\mathbf{x}^d, y) - 1 + \Phi_K^d\left(-\log(\rho)'(x_1^d)(y - x_1^d)\right) - \mathbf{q}_{\mathbf{x}^d}^d\left(-\log(\rho)'(x_1^d)(y - x_1^d)\right)\right).$$

Part (v) follows from this representation of $\frac{\partial^2}{\partial y^2}\beta(\mathbf{x}^d, y)$ and Lemma 11. □

Proof of Proposition 7. Fix an arbitrary $\mathbf{x}^d \in \mathcal{A}_d$. Let Z_1, W_1 be random variables, as in Proposition 22, that satisfy

$$\begin{aligned} f(Y_1^d) - f(x_1^d) &= f'(x_1^d)(Y_1^d - x_1^d) + \frac{f''(x_1^d)}{2}(Y_1^d - x_1^d)^2 + \frac{f'''(Z_1)}{6}(Y_1^d - x_1^d)^3, \\ \beta(\mathbf{x}^d, Y_1^d) &= \beta(\mathbf{x}^d, x_1^d) + \frac{\partial}{\partial y}\beta(\mathbf{x}^d, x_1^d)(Y_1^d - x_1^d) + \frac{\frac{\partial^2}{\partial y^2}\beta(\mathbf{x}^d, W_1)}{2}(Y_1^d - x_1^d)^2. \end{aligned}$$

Then, by the definition of $\tilde{\mathcal{G}}_d f(\mathbf{x}^d)$ in (14) and the fact $Y_1^d - x_1^d \sim N(0, l^2/d)$, we find

$$\begin{aligned} \tilde{\mathcal{G}}_d f(\mathbf{x}^d) &= \frac{l^2 f''(x_1^d)}{2}\beta(\mathbf{x}^d, x_1^d) + l^2 f'(x_j^d)\frac{\partial}{\partial y}\beta(\mathbf{x}^d, x_1^d) \\ &+ d\mathbb{E}_{Y_1^d}\left[\left(\beta(\mathbf{x}^d, x_1^d)\frac{f'''(Z_1)}{6} + f'(x_1^d)\frac{\frac{\partial^2}{\partial y^2}\beta(\mathbf{x}^d, W_1)}{2}\right)(Y_1^d - x_1^d)^3\right] \\ &+ d\mathbb{E}_{Y_1^d}\left[\left(\frac{f''(x_1^d)}{2}\frac{\frac{\partial^2}{\partial y^2}\beta(\mathbf{x}^d, W_1)}{2} + \frac{f'''(Z_1)}{6}\frac{\partial}{\partial y}\beta(\mathbf{x}^d, x_1^d)\right)(Y_1^d - x_1^d)^4\right] \\ &+ d\mathbb{E}_{Y_1^d}\left[\frac{f'''(Z_1)}{6}\frac{\frac{\partial^2}{\partial y^2}\beta(\mathbf{x}^d, W_1)}{2}(Y_1^d - x_1^d)^5\right]. \end{aligned}$$

By parts (ii) and (iv) in Lemma 12 and the definition of $\hat{\mathcal{G}}_d f(\mathbf{x}^d)$ in (17) we have $\hat{\mathcal{G}}_d f(\mathbf{x}^d) = \frac{l^2 f''(x_1^d)}{2}\beta(\mathbf{x}^d, x_1^d) + l^2 f'(x_j^d)\frac{\partial}{\partial y}\beta(\mathbf{x}^d, x_1^d)$. The three expectations in the display above can each be bounded by a constant times $\left(\sum_{i=1}^3 \|f^{(i)}\|_{\infty,1/2}\right) e^{|x_1^d|}d^{-1/2}$ using Proposition 22 and Lemma 12. For instance, the first expectation can be bounded above using (v) in Lemma 12:

$$\frac{d}{6}\mathbb{E}_{Y_1^d}\left[|f'''(Z_1)||Y_1^d - x_1^d|^3\right] + \frac{d(C_K+1)}{2}|f'(x_1^d)||\log(\rho'(x_1^d))|^2\mathbb{E}_{Y_1^d}\left[|Y_1^d - x_1^d|^3\right].$$

Proposition 22 yields $\frac{d}{6}\mathbb{E}_{Y_1^d}\left[|f'''(Z_1)||Y_1^d - x_1^d|^3\right] \leqslant C_0 e^{|x_1^d|}\|f'''\|_{\infty,1}d^{-1/2} \leqslant C_0 e^{|x_1^d|}\|f'''\|_{\infty,1/2}d^{-1/2}$ for some $C_0 > 0$. Moreover, $|f'(x_1^d)||\log(\rho'(x_1^d))|^2 \leqslant \|f'\|_{\infty,1/2}\|(\log(\rho)')^2\|_{\infty,1/2}e^{|x_1^d|}$ as $\log(\rho) \in \mathcal{S}^4$ and $f \in \mathcal{S}^3$. Hence $\frac{d(C_K+1)}{2}|f'(x_1^d)||\log(\rho'(x_1^d))|^2\mathbb{E}_{Y_1^d}\left[|Y_1^d - x_1^d|^3\right] \leqslant C_1 e^{|x_1^d|}\|f'\|_{\infty,1/2}d^{-1/2}$ for some $C_1 > 0$. Similarly, it follows that the second and third expectations above decay as d^{-1} and $d^{-3/2}$, respectively. This concludes the proof of the proposition. ☐

Lemma 13. *Recall that $\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)$ and $\sum_{i=2}^d K(x_i^d, Y_i^d)$ are defined in (18) and (16), respectively. Then there exists a constant C such that for all $d \in \mathbb{N}$ we have:*

$$\mathbb{E}_{\mathbf{Y}^d-}\left[\left|\sum_{i=2}^d K(x_i^d, Y_i^d) - \mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)\right|\right] \leqslant Cd^{-1/2} \quad \forall \mathbf{x}^d \in \mathcal{A}_d.$$

Proof. The difference in question is smaller than the sum of the following two terms:

$$\begin{aligned} T_3^d(\mathbf{x}^d) &= \mathbb{E}_{\mathbf{Y}^d-}\left[\left|\sum_{i=2}^d \frac{\log(\rho)''(x_i^d)}{2}\left((Y_i^d - x_i^d)^2 - \frac{l^2}{d}\right)\right|\right], \\ T_4^d(\mathbf{x}^d) &= \mathbb{E}_{\mathbf{Y}^d-}\left[\left|\sum_{i=2}^d \frac{\left(\log(\rho)'(x_i^d)\right)^3 1_A(x_i^d)}{3}(Y_i^d - x_i^d)^3\right|\right]. \end{aligned}$$

Note that $X_i := (Y_i^d - x_i^d)^2 - l^2/d$, $2 \leqslant i \leqslant d$, are zero mean IID with $\mathbb{E}[X_i^4] = 2l^4/d^2$. Hence, as $\log(\rho) \in \mathcal{S}^4$, we may apply Proposition 23 with the function $x \mapsto \log(\rho)''(x)$ and X_i,$2 \leqslant i \leqslant d$, to get $T_3^d(\mathbf{x}^d) \leqslant \|\log(\rho)''\|_{\infty,1/2}\left(2(l^4/d^2)\sum_{i=2}^d e^{|x_i^d|}\right)^{1/2} \leqslant C_0 d^{-1/2}$ for some constant $C_0 > 0$, where the second inequality follows from (10). Similarly, Proposition 23 and assumption (10), applied to the function $x \mapsto (\log(\rho)'(x))^3 1_A(x)$ and random variables $Y_i^d - x_i^d$, yield $T_4^d(\mathbf{x}^d) \leqslant C_1\left(\sum_{i=2}^d e^{|x_i^d|}/d^3\right)^{1/2} \leqslant C_2 d^{-1}$ for some constants $C_1, C_2 > 0$ and all $d \in \mathbb{N}$. ☐

Lemma 14. *There exist constants $c_1, c_1' > 0$, such that for any $d \in \mathbb{N}$, $i \in \{2, \cdots, d\}$, $\mathbf{x}^d \in \mathcal{A}_d$ and $x_i^d \notin A$, it holds*

$$\left|\log\varphi_i(t) - \left(i\frac{l^2}{2d}\log(\rho)''(x_i^d)t - \frac{l^2}{2d}\left(\log(\rho)'(x_i^d)\right)^2 t^2\right)\right| \leqslant \frac{c_1}{d^{3/2}}(t^2 + |t|^3), \quad |t| \leqslant c_1'\sqrt{d},$$

where $\varphi_i(t) := \mathbb{E}_{Y_i^d}[\exp(itK(x_i^d, Y_i^d))]$, $t \in \mathbb{R}$, is the CF of $K(x_i^d, Y_i^d)$ (cf. (16)).

Remark 5. Recall that the set A satisfies the conclusion of Proposition 4. The proof of Lemma 14 requires the control of the functions $\log(\rho)'$ and $\log(\rho)''$ on the complement of A, where they are unbounded. It is hence crucial that their argument x_i^d is the i-th coordinate of a point $\mathbf{x}^d \in \mathcal{A}_d$, since, through assumption (10), we have control over the size of x_i^d in terms of the dimension d of the chain. For an analogous reason we need $i > 1$. These facts plays a key role in the proof.

Proof. By Lemma 32, the following inequality holds for all $|t| \leqslant d/(4l^2|\log(\rho)''(x_i^d)|)$:

$$\left|\log \varphi_i(t) - \left(i\frac{l^2}{2d}\log(\rho)''(x_i^d)t - \frac{l^2}{2d}\left(\log(\rho)'(x_i^d)\right)^2 t^2\right)\right| \leqslant \frac{l^4}{d^2}\left(\frac{1}{2}\left(\log(\rho)''(x_i^d)\right)^2 t^2 + \left(\log(\rho)'(x_i^d)\right)^2 \left|\log(\rho)''(x_i^d)\right| |t|^3\right), \tag{23}$$

Since $\mathbf{x}^d \in \mathcal{A}_d$ and $i > 1$, assumption (10) implies that for any $f \in \mathcal{S}^0$ there exists $C_f > 0$ such that $|f(x_i^d)|^2/C_f \leqslant e^{|x_i^d|} \leqslant \sum_{i=2}^d e^{|x_i^d|} \leqslant 2d\rho(e^{|x|})$. Hence $|f(x_i^d)| \leqslant c_f\sqrt{d}$ for all $d \in \mathbb{N}$ and all $2 \leqslant i \leqslant d$, where $c_f := (2C_f\rho(e^{|x|}))^{1/2}$. Since $\log(\rho) \in \mathcal{S}^4$, both functions $f_1(x) := l^4(\log(\rho)''(x))^2/2$ and $f_2(x) := (\log(\rho)'(x))^2|\log(\rho)''(x)|$ are in $\mathcal{S}^0$. Then (23) and the constants $c_1 := \max\{c_{f_1}, c_{f_2}\}$ and $c_1' := 1/(4\sqrt{2c_{f_1}})$ yield the inequalities in the lemma. □

We now deal with the coordinates of $\mathcal{A}_d$ that are in A. Compared to Lemma 14, this is straightforward as it does not involve the remainder of the coordinates of the point in $\mathcal{A}_d$.

Lemma 15. *If $x \in A$, then $K(x, Y)$ (cf. (16)), where $Y \sim N(x, l^2/d)$, satisfies:*
(a) $\mu_K := \mathbb{E}_Y[K(x,Y)] \leqslant \frac{l^2c_A^2}{2d}$, where $c_A > 0$ is the constant in Proposition 4;
(b) $|\mathbb{E}_Y[(K(x,Y) - \mu_K)^2] - (\log(\rho)'(x))^2\frac{l^2}{d}| \leqslant C_1d^{-2}$ for some constant $C_1 > 0$ and all $d \in \mathbb{N}$;
(c) $\mathbb{E}_Y\left[|K(x,Y) - \mu_K|^3\right] \leqslant C_2d^{-3/2}$ for some constant $C_2 > 0$ and all $d \in \mathbb{N}$.
Moreover, the constants C_1 and C_2 do not depend on the choice of $x \in A$.

Proof. By definition of A in Proposition 4 we have $|\log(\rho)'(x)| \leqslant c_A$ and $|\log(\rho)''(x)| \leqslant c_A^2$ for $x \in A$. By (16), $\mu_K = \frac{l^2}{2d}\log(\rho)''(x)$ and (a) follows. Recall $\mathbb{E}_Y[(Y - x)^n]$ is either zero (if n is odd) or of order $d^{-n/2}$ (if n is even) and $\mathbb{E}_Y[(Y - x)^2] = l^2/d$. Hence the definition of K in (16), the fact $x \in A$ and part (a) imply the inequality in part (b). For part (c), note that an analogous argument yields $\mathbb{E}_Y\left[(K(x,Y) - \mu_K)^6\right] \leqslant C'd^{-3}$ for some constant $C' > 0$. Cauchy's inequality concludes the proof of the lemma. □

Lemma 16. *Let assumptions of Lemma 15 hold and denote by φ the characteristic function of $K(x, Y)$. There exist positive constants c_2 and c_2', such that the following holds for all $x \in A$:*

$$\left|\log \varphi(t) - \left(it\frac{l^2}{2d}\log(\rho)''(x) - \frac{t^2l^2}{2d}\log(\rho)'(x)^2\right)\right| \leqslant c_2\left(\frac{t^2}{d^2} + \frac{|t|^3}{d^{3/2}} + \frac{t^4}{d^2}\right), \quad |t| \leqslant c_2'\sqrt{d}. \tag{24}$$

Proof. Let $\sigma_K^2 := \mathbb{E}_Y[(K(x,Y) - \mu_K)^2]$ and recall $\mu_K = \frac{l^2}{2d}\log(\rho)''(x)$. By Lemma 31 we have

$$\left|\log \varphi(t) - \left(it\frac{l^2}{2d}\log(\rho)''(x) - \frac{t^2}{2}\sigma_K^2\right)\right| \leqslant |t|^3\mathbb{E}_Y|K(x,Y) - \mu_K|^3/6 + t^4\sigma_K^4/4, \quad |t| \leqslant \frac{1}{\sigma_K}. \tag{25}$$

By Lemma 15(b) we have $|\sigma_K^2 - l^2\log(\rho)'(x)^2/d| \leqslant C_1d^{-2}$. Hence $\sigma_K^2 \leqslant d^{-1}/\sqrt{c_2'}$, where $c_2' := 1/(l^2c_A^2 + C_1)^2$, and $\sigma_K^4 \leqslant C_1'd^{-2}$ for some $C_1' > 0$. This, together with Lemma 15(c), implies that there exists a constant $c_2 > 0$, such that the inequality in (24) follows from (25) for all $|t| \leqslant c_2'd^{1/2} \leqslant 1/\sigma_K$ and $x \in A$. □

Lemma 17. *For any $d \in \mathbb{N}$ and $\mathbf{x}^d \in \mathcal{A}_d$, let Φ_K^d and $\Phi_{\mathcal{N}}^d$ be the distribution functions of $\sum_{i=2}^d K(x_i^d, Y_i^d)$ and $\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)$. Then there exists $C > 0$ such that*

$$\sup_{x \in \mathbb{R}} \left| \Phi_{\mathcal{N}}^d(x) - \Phi_K^d(x) \right| \leqslant C d^{-1/2} \qquad \text{for every } d \in \mathbb{N} \text{ and } \mathbf{x}^d \in \mathcal{A}_d.$$

Proof. Let φ_K and $\varphi_{\mathcal{N}}$ be the CFs of $\sum_{i=2}^d K(x_i^d, Y_i^d)$ and $\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)$, respectively. We will compare φ_K and $\varphi_{\mathcal{N}}$ and apply Proposition 30 to establish the lemma. Let φ_i be the CF of $K(x_i^d, Y_i^d)$ and recall, by (18), $\varphi_{\mathcal{N}}(t) = \exp\left(\frac{1}{2}it\frac{l^2}{d}\sum_{i=2}^d \log(\rho)''(x_i^d) - \frac{1}{2}t^2\frac{l^2}{d}\sum_{i=2}^d \left(\log(\rho)'(x_i^d)\right)^2\right)$. Define the positive constants $c := \max\{c_1, c_2\}$ and $c' := \min\{1, l^2 J/(32c), c_1', c_2'\}$, where the constants c_1, c_1' (resp. c_2, c_2') are given in Lemma 14 (resp. Lemma 16) and J is as in assumption (12). Note that the constants c, c' do not depend on the choice of $\mathbf{x}^d \in \mathcal{A}_d$. Lemmas 14 and 16 imply the following inequality for all $d \in \mathbb{N}$ and $\mathbf{x}^d \in \mathcal{A}_d$:

$$|\log \varphi_K(t) - \log \varphi_{\mathcal{N}}(t)| \leqslant \sum_{i=2}^d \left| \log \varphi_i(t) - \left(it\frac{l^2}{2d}\log(\rho)''(x_i^d) - \frac{t^2 l^2}{2d}\left(\log(\rho)(x_i^d\right)^2\right)\right| \leqslant R(t),$$

for all $|t| \leqslant r$, where $r := c'\sqrt{d}$ and $R(t) := c(t^2 + |t|^3 + t^4/\sqrt{d})/\sqrt{d}$. Since $|t|^3 \leqslant \sqrt{d}c't^2$ and $t^4 \leqslant dc'^2 t^2$ for $|t| \leqslant r$, we have

$$R(t) \leqslant t^2(c/\sqrt{d} + cc' + cc'^2) \leqslant t^2(c/\sqrt{d} + 2cc') \qquad \text{for all } t \in [-r, r]. \tag{26}$$

By assumption (12), there exists $d_0' \in \mathbb{N}$ such that the variance $\sigma_{\mathcal{N}}^2(\mathbf{x}^d) = l^2/d\sum_{i=2}^d \left(\log(\rho)'(x_i^d)\right)^2$ of $\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)$ satisfies $\sigma_{\mathcal{N}}^2(\mathbf{x}^d) \geqslant l^2 J/2$ for all $d \geqslant d_0'$ and $\mathbf{x}^d \in \mathcal{A}_d$. Let $\gamma := 1/2$ and pick $d_0 \in \mathbb{N}$, greater than $\max\{d_0', (16cl^{-2}/J)^2\}$. Then, for any $d \geqslant d_0$, the inequality $c/\sqrt{d} \leqslant \gamma l^2 J/8$ holds. Since $c' \leqslant l^2 J/(32c)$, we have $2cc' \leqslant \gamma l^2 J/8$, and the bound in (26) implies $R(t) \leqslant \frac{1}{2}t^2\gamma l^2 J/2 \leqslant \frac{1}{2}t^2\gamma\sigma_{\mathcal{N}}^2(\mathbf{x}^d)$ for all $t \in [-r, r]$. By Proposition 30, for all $d \geqslant d_0$, $\sup_{x\in\mathbb{R}} \left|\Phi_{\mathcal{N}}^d(x) - \Phi_K^d(x)\right|$ is bounded above by

$$\int_{\mathbb{R}} \frac{R(t)}{\pi|t|} \exp\left(-\frac{(1-\gamma)\sigma_{\mathcal{N}}^2(\mathbf{x}^d)t^2}{2}\right) dt + \frac{12\sqrt{2}}{\pi^{3/2}\sigma_{\mathcal{N}}(\mathbf{x}^d) r} \leqslant C'/\sqrt{d},$$

where $C' := c\int_{\mathbb{R}}(|t| + t^2 + |t|^3)\exp(-l^2 J t^2/8)dt + \frac{24\sqrt{2}}{\pi^{3/2}l^2 J c'}$. Since the left-hand side of the inequality in the lemma is bounded above by 1, the inequality holds for all $d \in \mathbb{N}$ if we define $C := \max\{C', \sqrt{d_0}\}$. □

Proof of Proposition 8. Since $|1 \wedge e^y - 1 \wedge e^x| \leqslant |x - y|$ for all $x, y \in \mathbb{R}$, by Lemma 13 we have

$$\left| \mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)}\right] - \mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\sum_{i=2}^d K(x_i^d, Y_i^d)}\right]\right| \leqslant C' d^{-1/2}$$

for some constant $C' > 0$ and all $d \in \mathbb{N}$. Recall $e^x 1_{\{x<0\}} = 1 \wedge e^x - 1 + 1_{\{x\leqslant 0\}}$ for all $x \in \mathbb{R}\backslash\{0\}$. Hence Lemmas 13 and 17 yield

$$\begin{aligned} &\left| \mathbb{E}_{\mathbf{Y}^d-}\left[e^{\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)} 1_{\{\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)<0\}}\right] - \mathbb{E}_{\mathbf{Y}^d-}\left[e^{\sum_{i=2}^d K(x_i^d, Y_i^d)} 1_{\left\{\sum_{i=2}^d K(x_i^d, Y_i^d)<0\right\}}\right]\right| \\ &\quad \leqslant \left| \mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\mathcal{N}(\mathbf{x}^d, \mathbf{Y}^d)}\right] - \mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\sum_{i=2}^d K(x_i^d, Y_i^d)}\right]\right| + \left|\Phi_{\mathcal{N}}^d(0) - \Phi_K^d(0)\right| \leqslant C'' d^{-1/2} \end{aligned}$$

for some $C'' > 0$ and all $d \in \mathbb{N}$. The proposition follows. □

Proof of Proposition 9. For any $\mathbf{x}^d \in \mathcal{A}_d$, by [RGG97, Proposition 2.4], we have

$$\mathbb{E}_{\mathbf{Y}^d-}\left[e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}1_{\{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)<0\}}\right] = e^{\mu_{\mathcal{N}}(\mathbf{x}^d)+\frac{\sigma^2_{\mathcal{N}}(\mathbf{x}^d)}{2}}\Phi\left(-\sigma_{\mathcal{N}}(\mathbf{x}^d) - \frac{\mu_{\mathcal{N}}(\mathbf{x}^d)}{\sigma_{\mathcal{N}}(\mathbf{x}^d)}\right),$$
$$\mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}\right] = \mathbb{E}_{\mathbf{Y}^d-}\left[e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}1_{\{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)<0\}}\right] + \Phi\left(\frac{\mu_{\mathcal{N}}(\mathbf{x}^d)}{\sigma_{\mathcal{N}}(\mathbf{x}^d)}\right).$$

where Φ is the distribution of a standard normal random variable. Note first that it is sufficient to prove the inequality in the proposition for all $d > d_0$ for some $d_0 \in \mathbb{N}$, since the expectations above are bounded by 1 and we can hence increase the constant C so that the first d_0 inequalities are also satisfied.

Recall the formulas for $\mu_{\mathcal{N}}(\mathbf{x}^d)$ and $\sigma^2_{\mathcal{N}}(\mathbf{x}^d)$ from (18). By assumptions (12) and (13) it follows that $\left|\mu_{\mathcal{N}}(\mathbf{x}^d) + \sigma^2_{\mathcal{N}}(\mathbf{x}^d)/2\right| \leqslant ca_d/\sqrt{d}$ for some constant $c > 0$ and all large d and $\mathbf{x}^d \in \mathcal{A}_d$. Note that $S_a := \sup_{d\in\mathbb{N}}(a_d/\sqrt{d}) < \infty$ since $\{a_d\}_{d\in\mathbb{N}}$ is sluggish. The function $x \mapsto e^x$ is Lipschitz on $[-cS_a, cS_a]$ with constant e^{cS_a}. Consequently $\left|e^{\mu_{\mathcal{N}}(\mathbf{x}^d)+\sigma^2_{\mathcal{N}}(\mathbf{x}^d)/2} - 1\right| \leqslant e^{cS_a}a_d/\sqrt{d}$ for large d and uniformly in $\mathbf{x}^d \in \mathcal{A}_d$.

By assumption (12), for all large $d \in \mathbb{N}$ and all $\mathbf{x}^d \in \mathcal{A}_d$, we have $\sigma_{\mathcal{N}}(\mathbf{x}^d) \geqslant l\sqrt{J}/\sqrt{2}$. Hence, since the function $x \mapsto \sqrt{x}$ is Lipschitz with constant $c_1 := 1/(l\sqrt{2J})$ on $[\frac{l^2 J}{2}, \infty)$, we get $\left|\sigma_{\mathcal{N}}(\mathbf{x}^d)/2 - l\sqrt{J}/2\right| \leqslant (c_1/2)\left|\sigma^2_{\mathcal{N}}(\mathbf{x}^d) - l^2 J\right| \leqslant c_2 a_d/\sqrt{d}$, where constant $c_2 > 0$ exists by (12). Moreover, $|(\mu_{\mathcal{N}}(\mathbf{x}^d) + \sigma^2_{\mathcal{N}}(\mathbf{x}^d)/2)/\sigma_{\mathcal{N}}(\mathbf{x}^d)| \leqslant c_3 a_d/\sqrt{d}$ for $c_3 > 0$ and all large d.

Since $\sigma_{\mathcal{N}}(\mathbf{x}^d) + \mu_{\mathcal{N}}(\mathbf{x}^d)/\sigma_{\mathcal{N}}(\mathbf{x}^d) = \left(\mu_{\mathcal{N}}(\mathbf{x}^d) + \sigma^2_{\mathcal{N}}(\mathbf{x}^d)/2\right)/\sigma_{\mathcal{N}}(\mathbf{x}^d) + \sigma_{\mathcal{N}}(\mathbf{x}^d)/2$, the inequalities in the previous paragraph imply that there exists $c_4 > 0$ such that $|\sigma_{\mathcal{N}}(\mathbf{x}^d) + \mu_{\mathcal{N}}(\mathbf{x}^d)/\sigma_{\mathcal{N}}(\mathbf{x}^d) - l\sqrt{J}/2| \leqslant c_4 a_d/\sqrt{d}$ for large d and uniformly in $\mathbf{x}^d \in \mathcal{A}_d$. Since Φ is Lipschitz with constant $1/\sqrt{2\pi}$, there exists a constant $C_1' > 1$, such that

$$\left|\mathbb{E}_{\mathbf{Y}^d-}\left[e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}1_{\{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)<0\}}\right] - \Phi\left(\frac{-l\sqrt{J}}{2}\right)\right| \leqslant C_1'\frac{a_d}{\sqrt{d}}$$

holds for all large d and all $\mathbf{x}^d \in \mathcal{A}_d$. Similarly, $\left|\mathbb{E}_{\mathbf{Y}^d-}\left[1 \wedge e^{\mathcal{N}(\mathbf{x}^d,\mathbf{Y}^d)}\right] - 2\Phi\left(\frac{-l\sqrt{J}}{2}\right)\right| \leqslant C_2'\frac{a_d}{\sqrt{d}}$ for some $C_2' > 0$ all large d and all $\mathbf{x}^d \in \mathcal{A}_d$, and the proposition follows. □

4.3. **Proof of Proposition 2.** We will now prove the following result.

Proposition 18. *Let $a = \{a_d\}_{d\in\mathbb{N}}$ be a sluggish sequence and $p \in [1, \infty)$. There exists a constant C_4 (depending on a and p) such that for every $f \in \mathcal{S}^3$ and all $d \in \mathbb{N}$ we have:*

$$\|\mathcal{G}f - \mathcal{G}_d f\|_p \leqslant C_4\left(\sum_{i=1}^{3}\|f^{(i)}\|_{\infty,1/2}\right)\left(\frac{a_d}{\sqrt{d}} + e^{-a_d^2/p}\right).$$

In the case $p = 2$, define $a_d := \sqrt{2\log(d)}$ for $d \in \mathbb{N}\backslash\{1\}$ and note that Proposition 2 then follows as a special case of Proposition 18.

Lemma 19. *There exists a constant C such that for all $f \in \mathcal{S}^3$ and all $d \in \mathbb{N}$ we have:*

$$\max\left\{\left|\mathcal{G}f(\mathbf{x}^d)\right|, \left|\mathcal{G}_d f(\mathbf{x}^d)\right|\right\} \leqslant Ce^{|x_1^d|}\sum_{i=1}^{2}\|f^{(i)}\|_{\infty,1/2} \quad \forall \mathbf{x}^d \in \mathbb{R}^d.$$

Proof. The triangle inequality, definition (5) and $\log(\rho)'(x)f'(x) \leqslant \|\log(\rho)'\|_{\infty,1/2}\|f'\|_{\infty,1/2}e^{|x|}$ imply the bound in the lemma for $|\mathcal{G}f(\mathbf{x}^d)|$. To bound $|\mathcal{G}_d f(\mathbf{x}^d)|$, define

$$\tilde{\beta}(\mathbf{x}^d, y) := \mathbb{E}_{\mathbf{Y}^d_-}\left[1 \wedge \exp\left(\log(\rho)(y) - \log(\rho)(x_1^d) + \sum_{i=2}^d \log(\rho)(Y_i^d) - \log(\rho)(x_i^d)\right)\right]$$

for any $y \in \mathbb{R}$. Then, if q denotes the density of $Y_1^d - x_1^d \sim N(0, l^2/d)$, we get

$$\begin{aligned}\left|\mathcal{G}_d f(\mathbf{x}^d)\right| &= d\left|\mathbb{E}_{Y_1^d}\left[\left(f(Y_1^d) - f(x_1^d)\right)\tilde{\beta}(\mathbf{x}^d, Y_1^d)\right]\right| \\ &\leqslant d\int_0^\infty z\left|f'(w_1)\tilde{\beta}(\mathbf{x}^d, x_1^d + z) - f'(w_2)\tilde{\beta}(\mathbf{x}^d, x_1^d - z)\right| q(z)dz,\end{aligned} \tag{27}$$

where $w_1 \in (x_1^d, x_1^d + z)$ and $w_2 \in (x_1^d - z, x_1^d)$ satisfy $zf'(w_1) = f(x_1^d + z) - f(x_1^d)$ and $-zf'(w_2) = f(x_1^d - z) - f(x_1^d)$, respectively. Moreover, $|f'(w_1) - f'(w_2)| \leqslant 2z|f''(w_3)|$ holds for some w_3 in the interval $(x_1^d - z, x_1^d + z)$. Since $x \mapsto 1 \wedge e^x$ is Lipschitz with constant 1, we get

$$\left|\tilde{\beta}(\mathbf{x}^d, x_1^d + z) - \tilde{\beta}(\mathbf{x}^d, x_1^d - z)\right| \leqslant \left|\log(\rho)(x_1^d + z) - \log(\rho)(x_1^d - z)\right| \leqslant 2z|\log(\rho)'(w_4)|$$

for some $w_4 \in (x_1^d - z, x_1^d + z)$. By adding and subtracting $f'(w_2)\tilde{\beta}(\mathbf{x}^d, x_1^d + z)$ on the right-hand side of (27), applying the two bounds we just derived and noting that $\tilde{\beta} \leqslant 1$, we get

$$\left|\mathcal{G}_d f(\mathbf{x}^d)\right| \leqslant 2d\int_0^\infty z^2|f''(w_3)|q(z)dz + 2d\int_0^\infty z^2|f'(w_2)\log(\rho)'(w_4)|q(z)dz. \tag{28}$$

Note that, since $\max\{|w_3|, |w_2|, |w_4|\} \leqslant |x_1^d| + z$ and $\|f''\|_{\infty,1} \leqslant \|f''\|_{\infty,1/2}$, we have

$$|f''(w_3)| \leqslant \|f''\|_{\infty,1}e^{|w_3|} \leqslant \|f''\|_{\infty,1/2}e^{|x_1^d|+z}, \quad \log(\rho)'(w_4)f'(w_2) \leqslant \|\log(\rho)'\|_{\infty,1/2}\|f'\|_{\infty,1/2}e^{|x_1^d|+z},$$

which, together with inequality (28), implies the lemma. □

Proof of Proposition 18. By Theorem 3 (on $\mathcal{A}_d$) and Lemma 19 (on $\mathbb{R}^d \backslash \mathcal{A}_d$), there exists a constant $C > 0$ such that for any $f \in \mathcal{S}^3$ the following inequality holds:

$$\begin{aligned}\|\mathcal{G}_d f - \mathcal{G}f\|_p^p &= \int_{\mathcal{A}_d}\left|\mathcal{G}_d f(\mathbf{x}^d) - \mathcal{G}f(\mathbf{x}^d)\right|^p \rho_d(\mathbf{x}^d)d\mathbf{x}^d + \int_{\mathbb{R}^d\backslash\mathcal{A}_d}\left|\mathcal{G}_d f(\mathbf{x}^d) - \mathcal{G}f(\mathbf{x}^d)\right|^p \rho_d(\mathbf{x}^d)d\mathbf{x}^d \\ &\leqslant C\rho(e^{p|x|})\left(\frac{a_d^p}{d^{p/2}}\rho_d(\mathcal{A}_d) + \rho_d(\mathbb{R}^d\backslash\mathcal{A}_d)\right)\left(\sum_{i=1}^3 \|f^{(i)}\|_{\infty,1/2}\right)^p.\end{aligned}$$

Apply Proposition 5 and raise both sides of the inequality to the power $1/p$ to conclude the proof of the proposition. □

4.4. **Proof of Theorem 1.**

Lemma 20. *Assume that ρ is a strictly positive density in $\mathcal{C}^1$ and that* (4) *holds. Then, for any $d \in \mathbb{N}$, the RWM chain $\{\mathbf{X}_n^d\}_{n\in\mathbb{N}}$ is V-uniformly ergodic with $V := 1/\sqrt{\rho_d}$.*

Proof. The lemma follows from [JH00, Theorem 4.1] if we prove that the target ρ_d satisfies

$$\lim_{|\mathbf{x}^d|\to\infty} \frac{\mathbf{x}^d}{|\mathbf{x}^d|} \cdot \nabla \log(\rho_d(\mathbf{x}^d)) = \lim_{|\mathbf{x}^d|\to\infty} \sum_{i=1}^d \frac{x_i^d}{|x_i^d|}\log(\rho)'(x_i^d) = -\infty, \tag{29}$$

$$\liminf_{|\mathbf{x}^d|\to\infty} \mathbb{P}_{\mathbf{Y}^d}\left[\rho_d(\mathbf{Y}^d) \geqslant \rho_d(\mathbf{x}^d)\right] > 0. \tag{30}$$

Assumption (4) implies that the expression $x/|x| \cdot \log(\rho)'(x)$ is bounded above and takes arbitrarily large negative values as $|x| \to \infty$. This yields (29), since $|\mathbf{x}^d| \to \infty$ implies that $|x_i^d| \to \infty$ holds for at least one $i \in \{1, \dots, d\}$.

Condition (30) states that the acceptance probability in the RWM chain is bounded away from zero sufficiently far from the origin. To prove this, recall that $\mathbf{Y}^d \sim N(\mathbf{x}^d, l^2/d \cdot I_d)$ and define the set

$$B(\mathbf{x}^d) := \left\{\mathbf{y}^d \in \mathbb{R}^d \colon \frac{x_i^d}{|x_i^d|} \cdot (y_i^d - x_i^d) \in \left(\frac{-2l}{\sqrt{d}}, \frac{-l}{\sqrt{d}}\right) \text{ for all } i \leqslant d\right\},$$

where we interpret $x_i^d/|x_i^d| := 1$ if $x_i^d = 0$. Clearly $\inf_{\mathbf{x}^d\in\mathbb{R}^d} \mathbb{P}_{\mathbf{Y}^d}\left[B(\mathbf{x}^d)\right] > 0$. We now prove that if $|\mathbf{x}^d|$ is sufficiently large, then $\rho_d(\mathbf{y}^d) \geqslant \rho_d(\mathbf{x}^d)$ for all $\mathbf{y}^d \in B(\mathbf{x}^d)$, which implies (30).

By (4), far enough from zero, ρ is decreasing in a direction away from the origin. Therefore, there exists a compact interval $K \subset \mathbb{R}$ such that $(-2l/\sqrt{d}, 2l/\sqrt{d}) \subset K$ and $\rho(y) \geqslant \rho(x)$ whenever $x \notin K$ and $x/|x| \cdot (y - x) \in (-2l/\sqrt{d}, -l/\sqrt{d})$. We claim that for every $\mathbf{y}^d \in B(\mathbf{x}^d)$, the inequality $\rho(y_i^d)/\rho(x_i^d) \geqslant (\min_{x\in K} \rho(x))/(\max_{x\in\mathbb{R}} \rho(x)) \in (0, 1)$ holds. If $x_i^d \in K$, then $y_i^d \in K$ and the inequality follows trivially. If $x_i^d \notin K$, then, by the definition of K, we have $\rho(y_i^d)/\rho(x_i^d) \geqslant 1$. This proves the claim. Hence, for $\mathbf{y}^d \in B(\mathbf{x}^d)$ we have

$$\frac{\rho_d(\mathbf{y}^d)}{\rho_d(\mathbf{x}^d)} \geqslant \left(\max_{i\leqslant d} \frac{\rho(y_i^d)}{\rho(x_i^d)}\right) \cdot \left(\frac{\min_{x\in K} \rho(x)}{\max_{x\in\mathbb{R}} \rho(x)}\right)^{d-1}. \tag{31}$$

We now prove that the ratio $\rho(y_i^d)/\rho(x_i^d)$ takes arbitrarily large values as $|x_i^d| \to \infty$. To show this, pick $\mathbf{y}^d \in B(\mathbf{x}^d)$ and assume the inequality $y_i^d > x_i^d$. Then $x_i^d < 0$ and $y_i^d - x_i^d > l/\sqrt{d}$. Moreover the following holds

$$\frac{\rho(y_i^d)}{\rho(x_i^d)} = \exp\left(\log\left(\frac{\rho(y_i^d)}{\rho(x_i^d)}\right)\right) \geqslant 1 + \int_{x_i^d}^{y_i^d} \log(\rho)'(z)dz \geqslant 1 + l/\sqrt{d} \inf_{z<x_i^d+2l/\sqrt{d}} \log(\rho)'(z) \to \infty$$

as $x_i^d \to -\infty$ by (4). This, together with (31), implies (30). The case $y_i^d < x_i^d$ is analogous and the lemma follows. □

Proposition 21. *If a strictly positive ρ satisfies* (4) *and $\log(\rho) \in \mathcal{S}^{n_\rho}$ and $f \in \mathcal{S}^{n_f}$ for some integers $n_\rho, n_f \in \mathbb{N} \cup \{0\}$, then the function $\hat{f}$, defined in* (7)*, satisfies $\hat{f} \in \mathcal{S}^{\min(n_f+2, n_\rho+1)}$.*

Proof. Clearly, if $f \in \mathcal{C}^{n_f}$ and $\rho \in \mathcal{C}^{n_\rho}$ and if ρ is strictly positive, then $\hat{f} \in \mathcal{C}^{\min(n_f+2,n_\rho+1)}$. Pick $s > 0$. The L'Hospital's rule implies:

$$\lim_{x\to\infty} \frac{\hat{f}(x)}{e^{s|x|}} = \frac{2}{sh(l)} \lim_{x\to\infty} \frac{\int_{-\infty}^{x} \rho(y)(\rho(f) - f(y))dy}{e^{sx}\rho(x)} = \frac{2}{sh(l)} \lim_{x\to\infty} \frac{\rho(f) - f(x)}{se^{sx} + e^{sx}(\log(\rho))'(x)}.$$

The last limit is zero by (4). An analogous argument shows $\lim_{x\to-\infty} \hat{f}(x)/e^{s|x|} = 0$. Hence $\|\hat{f}\|_{\infty,s} < \infty$ holds for all $s > 0$. Since $h(l)\hat{f}'(x)/2 = \left(\int_{-\infty}^{x} \rho(y)(\rho(f) - f(y))dy\right)/\rho(x)$, this argument implies that $\|\hat{f}'\|_{\infty,s} < \infty$ holds for all $s > 0$. Hence $\hat{f} \in \mathcal{S}^1$.

Proceed by induction: assume that for all $k \leqslant n$ (where $1 \leqslant n < \min(n_f + 2, n_\rho + 1)$) we have $\|\hat{f}^{(k)}\|_{\infty,s} < \infty$ for any $s > 0$. Pick an arbitrary $u > 0$. By differentiating (6) we obtain

$$\hat{f}^{(n+1)} = -\sum_{k=0}^{n-1} \binom{n-1}{k} (\log(\rho))^{(k+1)} \hat{f}^{(n-k)} + \frac{2}{h(l)} (\rho(f) - f)^{(n-1)}.$$

Since $n \leqslant \min(n_\rho, n_f + 1)$, the induction hypothesis implies $\|\hat{f}^{(k)}\|_{\infty,u/2} < \infty$ for all $1 \leqslant k \leqslant n$. By assumption we have $\|f^{(n-1)}\|_{\infty,u} < \infty$ and $\|(\log(\rho))^{(k)}\|_{\infty,u/2} < \infty$ for all $1 \leqslant k \leqslant n$. Hence $\|\hat{f}^{(n+1)}\|_{\infty,u} < \infty$ holds for an arbitrary $u > 0$ and the proposition follows. □

Proof of Theorem 1. By Lemma 20, the RWM chain $\mathbf{X}^d$ with the transition kernel P_d is V-uniformly ergodic with $V = \rho_d^{-1/2}$. Moreover, by [RR97][Prop. 2.1 and Thm 2.1], P_d defines a self-adjoint operator on $\{g \in L^2(\rho_d) \colon \rho_d(g) = 0\}$ with norm $\lambda_d < 1$. Proposition 21 implies $\hat{f} \in \mathcal{S}^3$, since by assumption we have $f \in \mathcal{S}^1$ and $\log(\rho) \in \mathcal{S}^4$. By Remark 6(c) in Section 5 below we have $\hat{f}^2 \in \mathcal{S}^3$. Since $P_d \hat{f} = (1/d)\mathcal{G}_d \hat{f} + \hat{f}$, Lemma 19 implies that $(P_d \hat{f})^2(\mathbf{x}^d) \leqslant C_{\hat{f}} e^{2|x_1^d|}$ for some positive constant $C_{\hat{f}}$ and all $\mathbf{x}^d \in \mathbb{R}^d$. Hence (4) and the definition of V imply the inequality $\max\{\hat{f}^2, (P_d\hat{f})^2\} \leqslant cV$ for some constant $c > 0$. Consequently, by [MT09, Theorem 17.0.1], the CLT for the chain $\mathbf{X}^d$ and function $f + dP_d\hat{f} - d\hat{f}$ holds with some asymptotic variance $\hat{\sigma}^2_{f,d}$.

By [KV86, Gey92] we can represent $\hat{\sigma}^2_{f,d}$ in terms of a positive spectral measure $E_d(d\lambda)$ associated with the function $f - \rho(f) + dP_d\hat{f} - d\hat{f} = \mathcal{G}_d\hat{f} - \mathcal{G}\hat{f}$ as

$$\hat{\sigma}^2_{f,d} = \int_{\Lambda_d} \frac{1+\lambda}{1-\lambda} E_d(d\lambda),$$

where $\Lambda_d \subset [-\lambda_d, \lambda_d]$ denotes the spectrum of the self-adjoint operator P_d acting on the Hilbert space $\{g \in L^2(\rho_d) \colon \rho_d(g) = 0\}$. By the definition of the spectral measure $E_d(d\lambda)$ we obtain We can bound

$$\hat{\sigma}^2_{f,d} \leqslant \frac{1+\lambda_d}{1-\lambda_d} \int_{\Lambda_d} E_d(d\lambda) = \frac{1+\lambda_d}{1-\lambda_d} \|P_d(d\hat{f}) - d\hat{f} + f - \rho(f)\|_2^2 \leqslant \frac{2}{1-\lambda_d} \|\mathcal{G}_d\hat{f} - \mathcal{G}\hat{f}\|_2^2.$$

Finally, the result follows by Proposition 2. □

5. Technical results

The results in Section 5 use the ideas of Berry-Esseen theory and large deviations as well as the optimal Young inequality, and do not depend on anything in this paper that precedes them.

5.1. **Bounds on the expectations of test functions.** We start with elementary observations.

Remark 6. Recall that $\mathcal{S}^n$, $n \in \mathbb{N} \cup \{0\}$, is defined in (3). The following statements hold.

(a) If $n \leqslant m$, then $\mathcal{S}^m \subset \mathcal{S}^n$.
(b) For $n \in \mathbb{N}$, $f \in \mathcal{S}^n$ if and only if $f' \in \mathcal{S}^{n-1}$.
(c) If $f \in \mathcal{S}^n$ and $g \in \mathcal{S}^m$ then $f + g, fg \in \mathcal{S}^{\min(n,m)}$.

Proposition 22. *Pick an arbitrary $n \in \mathbb{N}$. Assume $f \in \mathcal{S}^n$, $k \leqslant n$, $x \in \mathbb{R}$ and $Y \sim N(x, \sigma^2)$. Then there exists measurable Z satisfying $f^{(k)}(Z)(Y-x)^k/k! = f(Y) - \sum_{i=0}^{k-1} f^{(i)}(x)(Y-x)^i/i!$ and $|Z-x| < |Y-x|$. Furthermore there exists a constant $C > 0$ (depending on n) such that, for any $m \in \mathbb{N}$ and $s > 0$ we have*

$$\mathbb{E}_Y\left[\left|f^{(k)}(Z)\right|^m |Y-x|^n\right] \leqslant Ce^{s^2\sigma^2}\mathbb{E}_Y\left[|Y-x|^n\right]\|f^{(k)}\|_{\infty,s/m}^m e^{s|x|}.$$

Proof. A random variable Z, defined via the integral form of the remainder in Taylor's theorem, lies a.s. between Y and x, implying $|Z-x| < |Y-x|$. Cauchy's inequality yields

$$\mathbb{E}_Y\left[\left|f^{(k)}(Z)\right|^m |Y-x|^n\right]^2 \leqslant \mathbb{E}_Y\left[\left|f^{(k)}(Z)\right|^{2m}\right]\mathbb{E}_Y\left[|Y-x|^{2n}\right]. \tag{32}$$

Since $f \in \mathcal{S}^n \subset \mathcal{S}^k$, we have $\sup_{x\in\mathbb{R}}\left|f^{(k)}(x)\right|^{2m} e^{-2s|x|} = \|f^{(k)}\|_{\infty,s/m}^{2m} < \infty$. As $Y \sim N(x,\sigma^2)$, the equality $\mathbb{E}_Y\left[|Y-x|^{2n}\right] = C^2\mathbb{E}_Y\left[|Y-x|^n\right]^2$ holds, where $C := (2\sqrt{\pi}\Gamma((2n+1)/2))^{1/2}/\Gamma((n+1)/2)$ and $\Gamma(\cdot)$ is the Euler gamma function. Hence, by (32), we get

$$\mathbb{E}_Y\left[\left|f^{(k)}(Z)\right|^m |Y-x|^n\right] \leqslant \frac{C}{\sqrt{2}}\|f^{(k)}\|_{\infty,s/m}^m\sqrt{\mathbb{E}_Y\left[e^{2s|Z|}\right]}\mathbb{E}_Y\left[|Y-x|^n\right].$$

It remains to note $\mathbb{E}_Y e^{2s(|Z|-|x|)} \leqslant \mathbb{E}_Y e^{2s|Z-x|} \leqslant \mathbb{E}_Y e^{2s|Y-x|} \leqslant 2\mathbb{E}_Y e^{2s(Y-x)} = 2e^{2s^2\sigma^2}$. □

Proposition 23. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a measurable (not necessary continuous) function such that $\|f\|_{\infty,1/2} < \infty$. Fix $n \in \mathbb{N}$, $\mathbf{x}^d \in \mathbb{R}^d$ and let $X_1, X_2 \ldots, X_d$ be IID copies of X, satisfying $\mathbb{E}\left[X^n\right] = 0$ and $\mathbb{E}\left[X^{2n}\right] < \infty$. Then the following inequality holds:*

$$\left|\mathbb{E}\left[\sum_{i=1}^d f(x_i^d)X_i^n\right]\right| \leqslant \|f\|_{\infty,1/2}\left(\mathbb{E}[X^{2n}]\sum_{i=1}^d e^{|x_i^d|}\right)^{1/2}.$$

Remark 7. Note that the assumptions of Proposition 23 imply that, if X is a non-zero random variable, then $n \in \mathbb{N}$ has to be odd.

Proof. By Jensen's inequality, the fact that $\mathbb{E}[X] = 0$ and the assumption on f we get

$$\mathbb{E}\left[\sum_{i=1}^d f(x_i^d)X_i^n\right]^2 \leqslant \mathbb{E}\left[\left(\sum_{i=1}^d f(x_i^d)X_i^n\right)^2\right] = \sum_{i=1}^d (f(x_i^d))^2\mathbb{E}\left[X_i^{2n}\right] \leqslant \|f\|_{\infty,1/2}^2\mathbb{E}[X^{2n}]\sum_{i=1}^d e^{|x_i^d|}.$$

□

5.2. **Deviations of the sums of IID random variables.**

Proposition 24. *Let $f \in \mathcal{S}^0$ be such that $\rho(f) = 0$ and let $a = \{a_d\}_{d\in\mathbb{N}}$ be a sluggish sequence. If the random vector $(X_{1,d}, \ldots, X_{d,d})$ follows the density ρ_d for all $d \in \mathbb{N}$, then for every $t > 0$ the following inequality holds for all but finitely many $d \in \mathbb{N}$:*

$$\mathbb{P}_{\rho_d}\left[\left|\frac{1}{d-1}\sum_{i=2}^d f(X_{i,d})\right| \geqslant \frac{ta_d}{\sqrt{d}}\right] \leqslant \exp(-t^2a_d^2/(3\rho(f^2))).$$

Remark 8. Proposition 24 is an elementary consequence of a deeper underlying result, that the sequence of random variables $\{\sum_{i=1}^d f(X_{i,d})/(a_d\sqrt{d})\}_{d\in\mathbb{N}}$ satisfies a moderate deviation principle with a good rate function $t \mapsto t^2/(2\rho(f^2))$ and speed a_d^2 (see [EL03] for details). The key inequality needed in the proof of Proposition 24 is given in the next lemma.

Lemma 25. *Let assumptions of Proposition 24 hold. If $\rho(f^2) > 0$, then for every closed $F \subseteq \mathbb{R}$ the following holds:*

$$\limsup_{d\to\infty} a_d^{-2} \log \mathbb{P}_{\rho_d}\left[\sum_{i=1}^d f(X_{i,d})/(a_d\sqrt{d}) \in F\right] \leqslant -\inf\{x^2/(2\rho(f^2)); x \in F\}.$$

Proof. The moderate deviations results [EL03, Thm 2.2, Lem. 2.5, Rem. 2.6] yield a sufficient condition for the above inequality. More precisely, for $X \sim \rho$, we need to establish:

$$\limsup_{d\to\infty} a_d^{-2} \log\left(d \cdot \mathbb{P}_\rho\left[|f(X)| \geqslant a_d\sqrt{d}\right]\right) = -\infty. \tag{33}$$

Fix an arbitrary $m \in \mathbb{N}$. Since $f \in \mathcal{S}^0$, we have $|f(x)| \leqslant \|f\|_{\infty,1/m} e^{|x|/m}$ for every $x \in \mathbb{R}$. Consequently, for all large d, we get

$$\mathbb{P}_\rho\left[|f(X)| \geqslant a_d\sqrt{d}\right] \leqslant \mathbb{P}_\rho\left[\|f\|^m_{\infty,1/m} e^{|X|} \geqslant d^{m/2}\right] \leqslant \|f\|^m_{\infty,1/m} \rho\left(e^{|X|}\right) d^{-m/2}.$$

Since $\{a_d\}_{d\in\mathbb{N}}$ is sluggish, $\exists C_0 > 0$ such that $a_d^{-2}\log(\|f\|^m_{\infty,1/m}\rho(e^{|X|})) < C_0 < a_d^{-2}\log(d)$ for all large $d \in \mathbb{N}$. Hence

$$a_d^{-2}\log(d\cdot\mathbb{P}_\rho[|f(X)| \geqslant a_d\sqrt{d}]) \leqslant a_d^{-2}(\log(\|f\|^m_{\infty,1/m}\rho(e^{|X|})) - (m/2-1)\log(d)) < -C_0(m/2-2),$$

for all large $d \in \mathbb{N}$. Since m was arbitrary, (33) follows. □

Proof of Proposition 24. Note that the proposition holds if $\rho(f^2) = 0$. Assume now $\rho(f^2) > 0$ and fix an arbitrary $t > 0$. Note that since $\{a_d\}_{d\in\mathbb{N}}$ is sluggish, so is $\{a'_d\}_{d\in\mathbb{N}}$, $a'_d := a_{d+1}\sqrt{d/(d+1)}$. Apply Lemma 25 to $F = \mathbb{R}\backslash(-t,t)$ and $\{a'_d\}_{d\in\mathbb{N}}$ to get the following inequality

$$\mathbb{P}_{\rho_{d-1}}\left[\left|\sum_{i=1}^{d-1} f(X_{i,d-1})/(a'_{d-1}\sqrt{d-1})\right| \geqslant t\right] \leqslant \exp\left(-3(a'_{d-1})^2t^2/(8\rho(f^2))\right) \tag{34}$$

for all large enough $d \in \mathbb{N}$. Since $3(a'_{d-1})^2/4 \geqslant 2a_d^2/3$ for all but finitely many $d \in \mathbb{N}$, the right-hand side in (34) is bounded above by $\exp(-(a_d)^2t^2/(3\rho(f^2)))$. Recall $\rho_d(\mathbf{x}^d) = \rho_{d-1}(\mathbf{x}^{d-1})\rho(x_d^d)$ and $a'_{d-1}\sqrt{d-1} = a_d(d-1)/\sqrt{d}$. Hence the left-hand side in inequality (34) equals $\mathbb{P}_{\rho_d}[|\sum_{i=2}^d f(X_{i,d})/(d-1)| \geqslant ta_d/\sqrt{d}]$ and the proposition follows. □

The next result is based on a combinatorial argument. A special case of Proposition 26 was used in [RGG97].

Proposition 26. *Let $n \in \mathbb{N}$ and a measurable $f\colon \mathbb{R} \to \mathbb{R}$ satisfy $\rho(f) = 0$ and $\rho(f^{2n}) < \infty$. If the random vector $(X_{1,d}, \dots, X_{d,d})$ is distributed according to ρ_d, then there exists a constant C, independent of d, such that $\mathbb{P}_{\rho_d}\left[\left|\frac{1}{d-1}\sum_{i=2}^d f(X_{i,d})\right| \geqslant 1\right] \leqslant Cd^{-n}$.*

Remark 9. The constant C in Proposition 26 may depend on $n \in \mathbb{N}$ and the function f.

Proof. Fix $n \in \mathbb{N}$ and let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Markov's inequality and the Multinomial theorem yield:

$$\mathbb{P}_{\rho_d}\left[\left|\frac{1}{d-1}\sum_{i=2}^d f(X_{i,d})\right| \geqslant 1\right] = \mathbb{P}_{\rho_d}\left[\left|\frac{1}{d-1}\sum_{i=2}^d f(X_{i,d})\right|^{2n} \geqslant 1\right] \leqslant \mathbb{E}_{\rho_d}\left(\frac{1}{d-1}\sum_{i=2}^d f(X_{i,d})\right)^{2n}$$

$$= (d-1)^{-2n} \sum_{\substack{k_2+k_3+\cdots+k_d=2n \\ k_2,k_3\ldots,k_d \in \mathbb{N}_0\setminus\{1\}}} \binom{2n}{k_2,k_3,\ldots,k_d} \prod_{i=2}^d \mathbb{E}_\rho\left[f(X_{i,d})^{k_i}\right],$$

where last equality holds, because the expectation of any summand of the form $\prod_{i=2}^d f(X_{i,d})^{k_i}$ is zero if any of the indices $k_i = 1$ since ρ_d has a product structure and $\rho(f) = 0$. By Jensen's inequality, $\prod_{i=2}^d \mathbb{E}_\rho\left[f(X_{i,d})^{k_i}\right] \leqslant \prod_{i=2}^d \mathbb{E}_\rho\left[f(X_{i,d})^{2n}\right]^{k_i/2n} = \rho(f^{2n})^{\sum_{i=2}^d \frac{k_i}{2n}} = \rho(f^{2n})$, and hence

$$\mathbb{P}_{\rho_d}\left[\left|\frac{1}{d-1}\sum_{i=2}^d f(X_{i,d})\right| \geqslant 1\right] \leqslant (2n)! \cdot \rho(f^{2n})(d-1)^{-2n} \cdot |\mathcal{N}_d|, \tag{35}$$

where $|\mathcal{N}_d|$ stands for the cardinality of the set

$$\mathcal{N}_d := \left\{(k_2,k_3,\ldots,k_d) \in \mathbb{N}_0^{d-1}; \quad \sum_{i=2}^d k_d = 2n \text{ and } k_i \neq 1 \text{ for all } 2 \leqslant i \leqslant d\right\}.$$

Inequality (35) and the next Claim prove the proposition.

Claim. $|\mathcal{N}_d| \leqslant C'd^n$ for a constant C' independent of d.

Proof of Claim. Consider a function $\zeta\colon \mathbb{N}_0^{d-1} \to \mathbb{N}_0^{d-1}$, $\zeta(a_2,a_3,\ldots,a_d) := (2\lfloor\frac{a_2}{2}\rfloor, 2\lfloor\frac{a_3}{2}\rfloor, \ldots 2\lfloor\frac{a_d}{2}\rfloor)$, that rounds each entry down to the nearest even number. Every element in the image $\zeta(\mathcal{N}_d)$ is a $(d-1)$-tuple of non-negative even integers with sum at most $2n$. Recall the number of k-combinations with repetition, chosen from a set of $d-1$ objects, equals $\binom{k+d-2}{k}$. There exists $C'' > 0$, such that

$$\begin{aligned} |\zeta(\mathcal{N}_d)| &\leqslant \left|\left\{(k_2,k_3,\ldots,k_d) \in \mathbb{N}_0^{d-1}; \quad \sum_{i=2}^d k_d \leqslant n\right\}\right| \\ &= \sum_{k=0}^n \left|\left\{(k_2,k_3,\ldots,k_d) \in \mathbb{N}_0^{d-1}; \quad \sum_{i=2}^d k_d = k\right\}\right| = \sum_{k=0}^n \binom{k+d-2}{k} \leqslant C''d^n. \end{aligned}$$

Note that the pre-image of a singleton under ζ contains at most 2^n elements (i.e. $(d-1)$-tuples) of $\mathcal{N}_d$. Indeed, by the definition of $\mathcal{N}_d$, at most n coordinates of an element are not zero and each can either reduce by one or stay the same. Hence, for $C' := C''2^n$, we have $|\mathcal{N}_d| \leqslant 2^n \cdot |\zeta(\mathcal{N}_d)| \leqslant C'd^n$. ∎

5.3. **Bounds on the densities of certain random variables.** The key step in the proof of Proposition 27 below is the optimal Young's inequality: for $p, q \geqslant 1$ and $r \in [1,\infty]$, such that $1/p + 1/q = 1 + 1/r$, and functions $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$, their convolution $f * g$ satisfies the inequality

$$\|f * g\|_r \leqslant \frac{C_pC_q}{C_r}\|f\|_p\|g\|_q, \quad \text{where} \quad C_s := \begin{cases} \sqrt{\frac{s^{1/s}}{s'^{1/s'}}}, \text{ if } s \in (1,\infty) \text{ and } 1/s + 1/s' = 1, \\ 1, \text{ if } s \in \{1,\infty\}. \end{cases} \tag{36}$$

For $s \in [1,\infty)$, $\|\cdot\|_s$ is the usual norm on $L^s(\mathbb{R})$ and $\|\cdot\|_\infty$ denotes the essential supremum norm on $L^\infty(\mathbb{R})$. The proof of (36) for $r < \infty$ is given in [Bar98, Thm 1]. In the case $r = \infty$, we have $C_pC_q/C_r = 1$ and the inequality in (36) follows from the definition of the convolution, translation invariance of the Lebesgue measure and Hölder's inequality.

Proposition 27. *Let $X_1, X_2, \ldots, X_d$ be independent random variables, each X_i with a bounded density q_i. The density Q_d of the sum $\sum_{i=1}^d X_i$ satisfies $\|Q_d\|_\infty \leqslant c\max_{i\leqslant d}\|q_i\|_\infty/\sqrt{d}$ for some constant $c > 0$.*

Remark 10. The factor $d^{-1/2}$ in the inequality of Proposition 27 above comes from (36) and is crucial for the analysis in this paper. The standard Young's inequality for convolutions would only yield $\|Q_d\|_\infty \leqslant c\max_{i\leqslant d}\|q_i\|_\infty$, which gives insufficient control over Q_d.

Proof of Proposition 27. Since random variables X_i are independent, the density of their sum is a convolution of the respective densities, $Q_d = \ast_{i=1}^d q_i$. For all i and each $t > 1$ we have $q_i \in L^\infty(\mathbb{R}) \cap L^1(\mathbb{R}) \subset L^t(\mathbb{R})$. Moreover, the following inequality holds for every $k \leqslant d-1$:

$$\|Q_d\|_\infty = \left\|\ast_{i=1}^d q_i\right\|_\infty \leqslant \left(C_{\frac{d}{d-1}}\right)^k C_{\frac{d}{k}} \left(\prod_{i=1}^k \|q_i\|_{\frac{d}{d-1}}\right) \left\|\ast_{i=k+1}^d q_i\right\|_{\frac{d}{k}} \tag{37}$$

We prove (37) by induction on k. For $k = 1$, note that d and $\frac{d}{d-1}$ are Hölder conjugates, i.e. $1/d + 1/(d/(d-1)) = 1$ Hence (36) for $r = \infty$, $q = d$, $p = d/(d-1)$, $f = q_1$ and $g = \ast_{i=2}^d q_i$ implies $\|Q_d\|_\infty \leqslant \|q_1\|_{\frac{d}{d-1}} \left\|\ast_{i=2}^d q_i\right\|_d$ and $C_{\frac{d}{d-1}} = C_d^{-1}$. Now assume (37) holds for some $k \leqslant d-2$. Since $(d/(d-1))^{-1} + (d/(k+1))^{-1} = 1 + (d/k)^{-1}$, the inequality in (36) implies

$$\left\|\ast_{i=k+1}^d q_i\right\|_{\frac{d}{k}} \leqslant \frac{C_{\frac{d}{d-1}} C_{\frac{d}{k+1}}}{C_{\frac{d}{k}}} \|q_{k+1}\|_{\frac{d}{d-1}} \left\|\ast_{i=k+2}^d q_i\right\|_{\frac{d}{k+1}}.$$

This inequality and the induction hypothesis (i.e. (37) for k) implies (37) for $k+1$.

Since q_1 is a density, we have $\|q_i\|_1 = 1$. Hence we find $\|q_i\|_{\frac{d}{d-1}} \leqslant \|q_i\|_1^{\frac{d-1}{d}} \|q_i\|_\infty^{\frac{1}{d}} = \|q_i\|_\infty^{\frac{1}{d}}$ for each i, and in particular $\prod_{i=1}^d \|q_i\|_{\frac{d}{d-1}} \leqslant \max_{i\leqslant d}(\|q_i\|_\infty)$. By (37) for $k = d-1$ we get

$$\|Q_d\|_\infty \leqslant \left(C_{\frac{d}{d-1}}\right)^d \prod_{i=1}^d \|q_i\|_{\frac{d}{d-1}} \leqslant \max_{i\leqslant d}(\|q_i\|_\infty) \left(C_{\frac{d}{d-1}}\right)^d.$$

Since $\lim_{d\to\infty}\sqrt{d}\left(C_{\frac{d}{d-1}}\right)^d = \sqrt{e}$, there exists $c > 0$ such that $\left(C_{\frac{d}{d-1}}\right)^d \leqslant c/\sqrt{d}$ for all $d \in \mathbb{N}$. □

Polynomials of continuous random variables play an important role in the proofs of Section 4.

Proposition 28. *Let X be a continuous random variable and p a polynomial. Then the random variable $p(X)$ has a density.*

Proof. The set $B := p\left((p')^{-1}(\{0\})\right)$ has finitely many points. Moreover, p is locally invertible on $\mathbb{R}\backslash B$ by the inverse function theorem and the inverses are differentiable. Hence, for any $x \notin B$, the set $p^{-1}((-\infty, x])$ is a disjoint union of intervals with boundaries that depend smoothly on x. Since $\mathbb{P}[p(X) \leqslant x] = \mathbb{P}[X \in p^{-1}((-\infty, x])]$, the proposition follows. □

Proposition 29. *Let $N = N(\mu, \sigma^2)$ be a normal random variable and p a polynomial satisfying $\inf_{x\in\mathbb{R}} |p'(x)| \geqslant c_p$ for some constant $c_p > 0$. Then the random variable $p(N)$ has a probability density function $q_{p(N)}$, which satisfies $\|q_{p(N)}\|_\infty \leqslant (c_p\sigma\sqrt{2\pi})^{-1}$.*

Proof. Obviously, p is strictly monotonic and thus a bijection. Moreover, the distribution $\Phi_{p(N)}(\cdot)$ of $p(N)$ takes the form $\mathbb{P}\left[N \leqslant p^{-1}(\cdot)\right]$ or $\mathbb{P}\left[N > p^{-1}(\cdot)\right]$. Hence, for any $x \in \mathbb{R}$, the density $q_{p(N)}$ of $p(N)$ satisfies $q_{p(N)}(x) = q_N\left(p^{-1}(x)\right)\left|\left(p^{-1}\right)'(x)\right| = q_N\left(p^{-1}(x)\right)/\left|p'\left(p^{-1}(x)\right)\right| \leqslant 1/(c_p\sigma\sqrt{2\pi})$, as the density of N, q_N, is bounded above by $(\sigma\sqrt{2\pi})^{-1}$. □

5.4. **CFs and distributions of near normal random variables.**

Proposition 30. *Let N be a normal random variable with mean μ and variance σ^2 and X a continuous random variable. Denote with φ_X, φ_N and Φ_X, Φ_N the CFs and the distributions of X and N, respectively. Assume there exist constants $r > 0$, $\gamma \in (0,1)$ and a function $R\colon \mathbb{R} \to \mathbb{R}$ such that $|\log\varphi_X(t) - \log\varphi_N(t)| \leqslant R(t) \leqslant \gamma\sigma^2t^2/2$ holds on $|t| \leqslant r$. Then*

$$\sup_{x\in\mathbb{R}} |\Phi_N(x) - \Phi_X(x)| \leqslant \int_{-r}^{r} \frac{R(t)}{\pi|t|}\exp\left(-\frac{(1-\gamma)\sigma^2t^2}{2}\right)dt + \frac{12\sqrt{2}}{\pi^{3/2}\sigma r}.$$

Remark 11. The result is a direct consequence of the Smoothing theorem (see [Kol06, Theorem 2.5.2]) commonly used to prove Berry-Esseen-type bounds, that relate CFs and distribution functions of random variables.

Proof. The Smoothing theorem implies

$$\sup_{x\in\mathbb{R}} |\Phi_N(x) - \Phi_X(x)| \leqslant \int_{-r}^{r} |\varphi_N(t) - \varphi_X(t)|/(\pi|t|)dt + 24\sup_{x\in\mathbb{R}}|\Phi_N'(x)|/(\pi r).$$

Note that, for any $z \in \mathbb{C}$, it holds $|e^z - 1| \leqslant |z|e^{|z|}$. For $z := \log(\varphi_X(t)/\varphi_N(t))$, this implies

$$|\varphi_X(t) - \varphi_N(t)| \leqslant |\varphi_N(t)||\log\varphi_X(t) - \log\varphi_N(t)|\exp(|\log\varphi_X(t) - \log\varphi_N(t)|) \qquad \forall t \in \mathbb{R}.$$

The result follows from this inequality, $\sup_{x\in\mathbb{R}}|\Phi_N'(x)| = 1/(\sigma\sqrt{2\pi})$ and $|\varphi_N(t)| = e^{-\sigma^2t^2/2}$:

$$\int_{-r}^{r} \frac{|\varphi_N(t) - \varphi_X(t)|}{\pi|t|}dt \quad\leqslant\quad \int_{-r}^{r} |\varphi_N(t)|\frac{R(t)}{\pi|t|}e^{R(t)}dt \quad\leqslant\quad \int_{-r}^{r} \frac{R(t)}{\pi|t|}\exp\left(-\frac{(1-\gamma)\sigma^2t^2}{2}\right)dt.$$

□

Lemma 31. *Let X be random variable with finite mean μ, variance σ^2 and absolute third central moment $\kappa := \mathbb{E}\left[|X-\mu|^3\right]$. Then, the characteristic function φ_X of X satisfies:*

$$\left|\log\varphi_X(t) - \left(i\mu t - \frac{\sigma^2}{2}t^2\right)\right| \leqslant \frac{\kappa|t|^3}{6} + \frac{\sigma^4t^4}{4} \quad \forall t \in \left[-\frac{1}{\sigma}, \frac{1}{\sigma}\right].$$

Proof. The result can be established by combining the elementary bound

$$\left|\mathbb{E}\left[e^{i(X-\mu)t} - \sum_{k=0}^{n} \frac{(it)^n}{n!}(X-\mu)^n\right]\right| \leqslant \frac{|t|^{n+1}}{(n+1)!}\mathbb{E}\left[|X-\mu|^{n+1}\right] \quad \forall t \in \mathbb{R}$$

and the fact that $z \in \mathbb{C}$, $|z| \leqslant 1/2$ implies $|(\log(1+z) - z| \leqslant |z|^2$ (see [Wil91, p. 188] for both).

□

Lemma 32. *Let $N = N(0, \sigma^2)$ and let $u, v \in \mathbb{R}$. The random variable $uN + vN^2$ has a characteristic function that satisfies*

$$\left|\log \varphi_{uN+vN^2}(t) - \left(iv\sigma^2 t - \frac{u^2\sigma^2}{2}t^2\right)\right| \leqslant 2v^2\sigma^4 t^2 + 2u^2|v|\sigma^4|t|^3 \quad \forall t \in \left[-\frac{1}{4|v|\sigma^2}, \frac{1}{4|v|\sigma^2}\right].$$

Proof. The CF φ_{uN+vN^2} can be explicitly computed using standard complex analysis

$$\varphi_{uN+vN^2}(t) = \mathbb{E}\left[e^{i(uN+vN^2)t}\right] = \frac{1}{\sqrt{1-2iv\sigma^2 t}} \exp\left(-\frac{u^2\sigma^2 t^2}{2(1-2iv\sigma^2 t)}\right) \quad \forall t \in \mathbb{R}.$$

The rest can then be shown using the elementary inequalities: $z \in \mathbb{C}$, $|z| \leqslant 1/2$ implies $|(\log(1+z) - z| \leqslant |z|^2$ and $|1/(1-z) - 1| \leqslant 2|z|$. □

References

[AC99] R. Assaraf and M. Caffarel. Zero-variance principle for monte carlo algorithms. *Physical Review Letters*, 83(23):4682–4685, 1999.

[Bar98] F. Barthe. Optimal young's inequality and its converse: a simple proof. *Geometric & Functional Analysis GAFA*, 8(2):234–242, 1998.

[Béd07] M. Bédard. Weak convergence of Metropolis algorithms for non-I.I.D. target distributions. *Ann. App. Prob.*, 17(4):1222–1244, 2007.

[BGJM11] S. Brooks, A. Gelman, G.L. Jones, and X.-L. Meng, editors. *Handbook of Markov chain Monte Carlo.* Handbooks of Modern Stat. Methods. Chapman & Hall/CRC, Boca Raton, FL, 2011.

[BPR+13] A. Beskos, N. Pillai, G. Roberts, J.-M. Sanz-Serna, and A. Stuart. Optimal tuning of the hybrid monte carlo algorithm. *Bernoulli*, 19(5A):1501–1534, 2013.

[BR08] M. Bédard and J.S. Rosenthal. Optimal scaling of Metropolis algorithms: Heading toward general target distributions. *Canadian Journal of Statistics*, 36(4):483–503, 2008.

[BR17] J. Bierkens and G.O. Roberts. A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. App. Prob.*, 27(2):846–882, 2017.

[BRS09] A. Beskos, G.O. Roberts, and A. Stuart. Optimal scalings for local Metropolis–Hastings chains on nonproduct targets in high dimensions. *Ann. App. Prob*, 19(3):863–898, 2009.

[DK12] P. Dellaportas and I. Kontoyiannis. Control variates for estimation based on reversible markov chain monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):133–161, 2012.

[DKPR87] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, 1987.

[DLP16] A.B. Duncan, T. Lelièvre, and G.A. Pavliotis. Variance reduction using nonreversible Langevin samplers. *Journal of Statistical Physics*, 163:457, 2016.

[DRVZ16] A. Durmus, G.O. Roberts, G. Vilmart, and K. Zygalakis. Fast Langevin based algorithm for MCMC in high dimensions. 2016. To appear in *Ann. App. Prob.*

[EL03] P. Eichelsbacher and M. Löwe. Moderate deviations for i.i.d. random variables. *ESAIM Probab. Stat.*, 7:209–218, 2003.

[Gey92] C.J. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.

[Hen97] S.G. Henderson. *Variance Reduction Via an Approximating Markov Process.* PhD thesis, Department of Operations Research, Stanford University, 1997.

[JH00] S.F. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, 85(2):341–361, 2000.

[JLM15] B. Jourdain, T. Lelièvre, and B. Miasojedow. Optimal scaling for the transient phase of the random walk Metropolis algorithm: The mean-field limit. *Ann. App. Prob.*, 25(4):2263–2300, 2015.

[Kol06] J.E. Kolassa. *Series approximation methods in statistics*, volume 88. Springer Science & Business Media, 2006.

[KV86] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.

[Mey08] S. Meyn. *Control techniques for complex networks*. Cambridge University Press, Cambridge, 2008.

[MPS12] J.C. Mattingly, N.S. Pillai, and A. Stuart. Diffusion limits of the Random walk Metropolis algorithm in high dimensions. *Ann. App. Prob*, 22(3):881–930, 2012.

[MT09] S. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009.

[MV16] A. Mijatović and J. Vogrinc. On the Poisson equation for Metropolis-Hastings chains. *arXiv preprint arXiv:1511.07464*, 2016. To appear in *Bernoulli*.

[OGC17] C.J. Oates, M. Girolami, and N. Chopin. Control functionals for monte carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):695–718, 2017.

[PMG14] T. Papamarkou, A. Mira, and M. Girolami. Zero variance differential geometric markov chain monte carlo algorithms. *Bayesian Analysis*, 9(1):97–128, 2014.

[PST12] N.S. Pillai, A. Stuart, and A.H. Thiéry. Optimal scalings and diffusion limits for the Langevin algorithm in high dimensions. *Ann. App. Prob*, 22(6):2320–2356, 2012.

[RGG97] G.O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.

[RR97] G.O. Roberts and J.S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, 2:no. 2, 13–25 (electronic), 1997.

[RR98] G.O. Roberts and J.S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268, 1998.

[RR01] G.O. Roberts and J.S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical science*, 16(4):351–367, 2001.

[Tie94] L. Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1762, 1994. With discussion and a rejoinder by the author.

[Wil91] D. Williams. *Probability with martingales*. Cambridge university press, 1991.

Department of Statistics, University of Warwick, and The Alan Turing Institute, UK
Email address: `aleks.mijatovic@gmail.com`

Department of Statistics, University of Warwick, UK
Email address: `Jure.Vogrinc@warwick.ac,uk`